

**EDUCATION OVER THE LIFE-CYCLE:
CLASS SIZE EFFECTS, RETURNS TO EDUCATION,
AND WAGE TRAJECTORIES AFTER JOB LOSS**

Dissertation
submitted to the
Faculty of Business, Economics and Informatics
of the University of Zurich

to obtain the degree of
Doktor der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by

Simone Balestra
from Gerra Gambarogno, Ticino

approved in September 2015 at the request of

Prof. Dr. Uschi Backes-Gellner

Prof. Dr. Eric P. Bettinger

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, 16.09.2015

Chairman of the Doctoral Board: Prof. Dr. Josef Zweimüller

Acknowledgments

First and foremost, I would like to thank my doctoral advisor, Professor Uschi Backes-Gellner, for her outstanding supervision, guidance, excellent knowledge, and constant support throughout my graduate studies. I am not only indebted to her for encouraging me to start down this educational path, but also for contributing a great deal to my advancement. She gave me valuable advice both inside and outside of academic research.

I would like to express my gratitude to my doctoral co-advisor, Professor Eric Bettinger, who made my research stay at Stanford University possible. He contributed significantly to the improvement of my academic skills through his valuable comments and inspiring ideas. I will always be thankful for the time he invested in supporting my research.

My thanks go to colleagues at the University of Zurich and Stanford University for constructive discussions and the provision of a collegial working atmosphere. In particular, I would like to thank Dr. Simon Janssen and Dr. Christian Rupiotta for all the long-lasting conversations about our research and for their advice regarding many aspects of life.

I am also grateful to Professor Edward Lazear, Professor Carmit Segal, and Professor Stefan Wolter for their helpful feedback during our meetings and seminars. For editorial help and guidance, my thanks go to Natalie Reid and Lindsay Brown.

I gratefully acknowledge financial support from the Swiss National Science Foundation (SNSF) for my visiting scholarship at Stanford University and from the Swiss State Secretariat for Education, Research and Innovation (SERI) through its Leading House on the Economics of Education, Firm Behavior and Training Policies.

Last but not least, I would like to thank my family. Special thanks go to my siblings Giulia, Siro, and Martino for being always there for me. I am grateful to my aunts Silvia

and Stefania, who stood by me throughout this journey. Finally, I wish to thank my parents, Daniela and Brenno, for their endless support and motivation during all those years making it possible to pursue this education. I will always be grateful for their belief in me.

Contents

1	Introduction	1
2	Distributional Effects of Class Size and Teacher Aide	9
2.1	Introduction	9
2.2	Theory and Empirical Background	12
2.3	Data and Descriptive Statistics	15
2.3.1	Project STAR and Follow-up Data	15
2.3.2	Covariate Balance in STAR	17
2.3.3	Attrition Analysis	21
2.4	Econometric Models	24
2.4.1	The Education Production Function	24
2.4.2	Estimation of Distributional Effects	26
2.5	Results	32
2.5.1	Distributional Effects During Project STAR	32
2.5.2	Effects in Later Grades and High School	41
2.6	Conclusions and Discussion	45
2.7	Appendix	48

2.7.1	Descriptive Statistics over the Achievement Distribution	48
2.7.2	Robustness Checks	51
3	Returns to Education over the Wage Distribution	53
3.1	Introduction	53
3.2	Theory and Empirical Background	56
3.3	Data and Descriptive Statistics	61
3.4	Methods	63
3.4.1	The Wage Equations	65
3.4.2	Instrumental Variable Quantile Regression	66
3.4.3	Identification Strategy	69
3.5	Results and Discussion	72
3.5.1	Causal Returns to Education over Wage Distribution	72
3.5.2	Heterogeneous Returns Between and Within Types of Education .	77
3.6	Conclusions	81
3.7	Appendix	85
3.7.1	The Swiss Education System	85
3.7.2	Analytic Sample	86
3.7.3	Summary of the Reform of 1970	87
3.7.4	Alternative Instrument and Over-identified Models	89
3.7.5	Returns to Vocational and Academic Education Instrumented . .	92
4	The Effects of Involuntary Separations on Wage Trajectories	93
4.1	Introduction	93

4.2	Theory and Empirical Background	96
4.3	Data and Descriptive Statistics	99
4.4	Methods	100
4.4.1	The Wage Equation	100
4.4.2	Estimation Strategy	102
4.5	Empirical Findings	105
4.5.1	Effect of Involuntary Separation on Annual Income and Hourly Wage	105
4.5.2	Plausibility Checks	110
4.5.3	The Determinants of Involuntary Separations	113
4.6	Conclusions	116
4.7	Appendix	119
4.7.1	PPML without Zeros	119
4.7.2	Additional Robustness Checks	120
4.7.3	Linear Probability Models	122
5	Final Remarks	123
	References	130
	Curriculum Vitae	148

List of Tables

2.3.1 DESCRIPTIVE STATISTICS AND COVARIATE BALANCE	19
2.3.2 DIFFERENTIAL ATTRITION IN STAR AND IN FOLLOW-UP SURVEYS	22
2.5.1 CLASS SIZE EFFECTS IN KINDERGARTEN AND FIRST GRADE	34
2.5.2 CLASS SIZE EFFECTS IN KINDERGARTEN AND FIRST GRADE, SUB-SAMPLE .	38
2.5.3 CLASS SIZE EFFECTS IN FOURTH AND EIGHTH GRADES	42
2.5.4 CLASS SIZE EFFECTS IN HIGH SCHOOL	44
2.7.1 DESCRIPTIVE STATISTICS AND COVARIATE BALANCE OVER THE DISTRIBUTION (BELOW THE MEDIAN)	49
2.7.2 DESCRIPTIVE STATISTICS AND COVARIATE BALANCE OVER THE DISTRIBUTION (ABOVE THE MEDIAN)	50
2.7.3 ROBUSTNESS CHECKS, Z-SCORES AND IMPUTED SCORES	52
3.3.1 DESCRIPTIVE STATISTICS	63
3.3.2 DESCRIPTIVE STATISTICS OVER WAGE DISTRIBUTION	64
3.5.1 RETURNS TO EDUCATION, OLS AND QR ESTIMATES	73
3.5.2 RETURNS TO EDUCATION, TSLS ESTIMATES	74
3.5.3 RETURNS TO EDUCATION, TSLS AND IVQR ESTIMATES	76
3.5.4 RETURNS TO VOCATIONAL AND ACADEMIC EDUCATION, OLS AND QR . . .	79

3.7.1 SAMPLE CONSTRUCTION	86
3.7.2 COMPULSORY EDUCATION EXPANSION	87
3.7.3 CHANGES IN SCHOOL-YEAR START	88
3.7.4 RETURNS TO EDUCATION, TSLS ESTIMATES	90
3.7.5 OVER-IDENTIFIED IVQR	91
3.7.6 RETURNS TO VOCATIONAL AND ACADEMIC EDUCATION, TSLS AND TWO- STAGE QR	92
4.3.1 DESCRIPTIVE STATISTICS	101
4.5.1 EARNING EFFECTS OF INVOLUNTARY SEPARATIONS, FE ESTIMATION	106
4.5.2 EARNING EFFECTS OF INVOLUNTARY SEPARATIONS, PPML ESTIMATION . .	108
4.5.3 WAGE EFFECTS FOR DIFFERENT TYPES OF JOB LOSS, FE ESTIMATION . .	111
4.5.4 WAGE EFFECTS FOR DIFFERENT TYPES OF JOB LOSS, PPML ESTIMATION	112
4.5.5 DETERMINANTS OF INVOLUNTARY SEPARATION, PROBIT MODELS	115
4.7.1 EARNING EFFECTS OF INVOLUNTARY SEPARATIONS, PPML WITHOUT ZEROS	119
4.7.2 INCOME EFFECTS FOR DIFFERENT TYPES OF JOB LOSS, FE ESTIMATION .	120
4.7.3 INCOME EFFECTS FOR DIFFERENT TYPES OF JOB LOSS, PPML ESTIMATION	121
4.7.4 DETERMINANTS OF INVOLUNTARY SEPARATION, LPM	122

List of Figures

2.5.1 Distributional Effect of Small Class, Kindergarten	33
2.5.2 Distributional Effect of Small Class, First Grade	35
2.5.3 Distributional Effect of Aide, Kindergarten	36
2.5.4 Distributional Effect of Aide, First Grade	37
2.5.5 Distributional Effect of Small and Aide on Test Scores, Sub-Samples . . .	39
3.5.1 Returns to Education, Quantile Regression Estimates	73
3.5.2 Returns to Education, IVQR Estimates	77
3.5.3 Returns to Vocational and Academic Education, Quantile Regression Es- timates	79
3.5.4 Academic Education Premium, Quantile Regression Estimates	80
3.7.1 The Swiss Education System (Source: Swiss Federal Statistical Office) . .	85

Chapter 1

Introduction

*The direction in which education starts a man
will determine his future life.*

Plato—The Republic (Book IV)

People spend most of their youth in education. On average, students in OECD countries will receive about eight thousand hours of instruction during their primary and lower secondary education, and 80 percent of these students will continue to upper secondary or tertiary education (OECD, 2013). As the amount of human capital available in the labor force is a crucial element for a country's growth and well-being (Goldin and Katz, 2009), education is important for both the present and the future. Individuals thus have strong incentives for pursuing more education, and governments have incentives for building on the skills of the population through education.

Since the pioneering theoretical contributions of Becker (1962) and Schultz (1961) and the early empirical work of Griliches (1977) and Mincer (1974), research in economics of education has been examining the effects and outcomes of individual investments in education. The economic literature shows three main outcomes of education: knowledge,

earnings, and employability. First, education empowers individuals by increasing their knowledge, thereby helping people make competent decisions about their life inside—and outside—the labor market (Hanushek, 2013). Second, income from employment tends to rise in line with an individual’s level of education (Harmon et al., 2003). In OECD countries, for example, workers with tertiary education earn on average 1.5 times as much as workers with only an upper secondary education (OECD, 2012). Third, educational attainment has a large impact on employability. Individuals with more education are more likely to have a job and to be working full-time than those without (Kettunen, 1997).

The aim of this dissertation is to provide an elaborate analysis of the impact of education on these three outcomes, with a particular focus on causal relationships and heterogeneous effects. The first project of this dissertation investigates the effects of early childhood educational practices on student knowledge, measured by standardized test scores. Using data from a randomized experiment, I estimate the effect of smaller classes and classes with a teacher’s aide on pupil test scores. I examine effect heterogeneity for pupils of different achievement levels and then study the persistence of these effects in later grades. In the second project, I focus on the second educational outcome of interest, earnings. I examine the monetary returns to education, with a particular focus on the effects of investing in education over the distribution of wages, to uncover potential heterogeneous effects of education on wages, and to discover who would profit most from acquiring additional schooling. The third project of the dissertation studies the third outcome deriving from investments in human capital: employability. I investigate size and persistence of wage losses after lay-off, to discover whether individuals undergo permanent scars when they lose their job. Furthermore, I examine the determinant of involuntary job losses to understand how the likelihood of experiencing an involuntary job loss is moderated by quantity and type of education.

The potential of education is limited if an individual’s cognitive skills are not de-

veloped early (Doyle et al., 2009), especially because essential competencies are better acquired during the first years of life (Doyle et al., 2013). Therefore, high-quality kindergarten and primary school lay a strong learning foundation that is fundamental to the rest of the lives of the individuals involved. Class size is probably the most used—and universally accepted—measure for education quality (Hanushek, 1999), and several developed countries are applying policies of class size reduction with the aim of improving student achievement (Lazear, 2001). However, the empirical evidence on the topic is often mixed (Fredriksson et al., 2014), and the long-term effects of lower pupil-to-teacher ratios are also being debated (Chetty et al., 2011).

To fill these gaps, in chapter two I use data from a large-scale student-to-teacher ratio experiment to estimate the impact on achievement of being in a smaller class or a regular-size class with a teacher’s aide, compared to a regular-size one. Given that children have different levels of ability, I let the effect of pupil-to-teacher ratio be heterogeneous over the achievement distribution. This distributional analysis allows me to answer the question of who profits more from being in a smaller class or in a regular-size class with a teacher’s aide. Results show that mid-achievers profit the most from being assigned to a smaller class. Students at the bottom or top of the achievement distribution experience only minimal gains from being in a small class.

The analysis also reveals a positive and significant effect of a regular class with a teacher’s aide for students at the bottom of the achievement distribution, an effect stronger for boys and disadvantaged children (low socioeconomic status and ethnic minorities)⁰. In line with previous research (Krueger and Whitmore, 2001), I show that being assigned to a small class during early grades has a positive effect on test scores in later grades, on the likelihood of on-time high school graduation, and on taking college readiness assessment exams. Interestingly, the positive effects on test scores are driven mostly by high achievers, suggesting that the long-term benefits of being in a small class shrink more quickly for low- and mid-achieving students. By contrast, the positive effects

on on-time high school graduation and on college exam-taking are present for the entire population of students.

After focusing on the impact on achievement of lower pupil-to-teacher ratios in early education, chapters three and four examine the value of education at a later stage of life, i.e., in the labor market. Although the literature reports sizable returns to education (Hartog and Oosterbeek, 1998), there are at least three considerations that need further research and discussion. First, whether a positive association between educational attainment and wage implies a causal relationship remains unclear.¹ Although individuals with higher ability and motivation acquire more education and have higher wages, such individuals may have earned the same wage even without education. Second, assuming the return to education to be equal for every individual appears unrealistic. In David Card's words,² "Is the labor force reasonably well-described by a constant return to education for all workers?" As the answer is "probably not," we should expect heterogeneous returns to education, especially for individuals at different wage levels. Third, most of the research focuses on the pure monetary benefits of education. However, human capital externalities (Acemoglu and Angrist, 2001) such as employability might also exist, and are likely to be heterogeneous as well (Oreopoulos and Salvanes, 2011). Employability in particular plays a crucial role when workers are laid off, because workers with different educational backgrounds might have completely different re-employment prospects (Farber, 2003).

Taken together, these three considerations make the analysis of investments in human capital difficult yet compelling. In this dissertation, I take these considerations into account by estimating heterogeneous returns to education (chapter three) and both the effects and determinants of involuntary separations (chapter four). In chapter three, I use administrative and survey data from Switzerland to identify the causal link between education and wages at different points of the distribution of wages. The analysis allows

¹See, e.g., Ashenfelter and Krueger (1994); Card (2012); Griliches and Mason (1972); Spence (1973); Willis and Rosen (1979).

²See Card (1994), p. 33.

me to discover the potential heterogeneous effects of education on wages, answering the question of who would profit the most from additional education. Building on a theoretical model of endogenous schooling with heterogeneous returns,³ my empirical results show that there is no unique causal effect of education and that for each individual the effect may deviate from those extensively documented by mean regression or instrumental variable estimation. In particular, once both the endogeneity of schooling and the heterogeneity in returns are taken into account, I estimate higher returns in lower parts of the wage distribution. Interpreting the wage distribution as a proxy for (unobserved) ability, the results suggest that higher-ability individuals have higher wages, but the slope of their wage-education profile is flatter than that for lower-ability individuals.

Observing a steeper slope at the bottom end of the distribution has important implications for both research and policy. From the research perspective, the finding indicates that more able individuals acquire more schooling because they face lower marginal costs and not because they experience higher marginal benefits. This gives empirical support to early theoretical contributions by Ashenfelter and Rouse (1998). From a policy perspective, it suggests that public policy should focus on the lower part of the distribution to efficiently allocate resources. Moreover, similar to the findings of chapter two, my analysis reveals that enhancing educational outcomes of individuals at different positions in the ability distribution not only calls for different practices (e.g., having a teacher's aide) but also implies different returns to human capital investments. Although this might appear rather intuitive, my dissertation is the first in providing clear and causal empirical evidence on these heterogeneous effects of education.

In a further step, I analyze another type of heterogeneity by comparing the returns to one extra year of academic education with the returns to one extra year of vocational education, to investigate whether one track brings a return premium at any point in the wage distribution. Looking at the heterogeneity patterns, I observe two relevant features

³Original model by Card (1994), further extended by Ashenfelter and Rouse (1998) and Arias et al. (2001).

of the returns to vocational and academic education. First, I find significant heterogeneity within each educational path, and, second, a comparison between the two tracks reveals that academic education does not always yield higher returns. In the upper part of the wage distribution, workers with an academic background experience higher returns than individuals with a vocational background. However, in lower parts of the distribution, vocational education brings higher returns than academic education. These results imply that answering the question of who would profit more from investments in human capital is not as easy as descriptive statistics or mean regression might suggest. Indeed, the answer depends on the individual's position in the wage (ability) distribution.

Chapter four studies the last outcome of investments in human capital I introduced before, i.e., employability. Using data from the Swiss Labor Force Survey, I estimate the earning losses of workers experiencing an involuntary job separation. To avoid selection biases arising from low-productive workers being dismissed (Akerlof, 1970; Gibbons and Katz, 1991), I apply several econometric specifications to achieve valid estimates of the earning losses. The results show large and persistent wage and productivity losses following an involuntary job loss, compared with a worker's expected level had the separation never happened. Analyzing other reasons for separation, I observe that the earning loss pattern is unique for involuntary separations, because no other type of separation implies such long-term losses.

Given the size and persistence of earnings losses, I also study the determinants of involuntary separations, with a particular focus on education as a protection against job loss. Previous literature (Kettunen, 1997; Mincer, 1991) suggests that the incidence of unemployment is lower among highly educated workers. Going beyond these findings, I focus not only on education level but also on the type of education, distinguishing between academic and vocational tracks. I find that tertiary education plays a major role in reducing the risk of job loss, independent of the type. This result suggests that both tertiary academic and tertiary vocational education are the best protection against invol-

untary separation. Given the impossibility of avoiding persistent losses once separation occurs, the findings of chapter four highlight the importance of reducing the likelihood of involuntary separation before it happens. Investments in human capital appear to help the most in preventing involuntary job losses, implying that outcomes depend not only on the reason for job loss but also on education.

Chapter five synthesizes the key elements of the preceding chapters and concludes. It also presents and discusses further research questions and policy implications that follow from the findings of this dissertation.

Chapter 2

Distributional Effects of Class Size and Teacher Aide

[A version of this paper is also available as: “Revisiting class-size effects: where they come from and how long they last” Swiss Leading House Working Paper no.102]

2.1 Introduction

In the last two decades, the subject that has received the most public, political, and academic attention in education economics is class size (Hoxby, 2000). Reducing class size to increase student achievement is a policy measure that has gained major consideration in the U.S., Europe, and Australia. Currently, many countries have enacted or are considering class size reduction with the aim of improving student achievement (Müller, 2013). In the U.S., for example, the student-to-teacher ratio in a given district is frequently used as a measure for education quality, and comparisons across districts are used as indices

of equity (Hanushek, 1999). However, there is still debate on the effects of class size, and we contribute to the literature by analyzing whether these effects are systematically different for students at different achievement levels.

Although smaller classes should theoretically have positive effects on student achievement (Lazear, 2001; Todd and Wolpin, 2003), empirical evidence is often mixed. Studies that rely on experimental data usually find positive effects of class size reduction on student test scores (Fletcher, 2009; Krueger, 1999). In contrast, most non-experimental studies report small or negligible class-size effects (Angrist and Lavy, 1999; Hoxby, 2000; Woessmann and West, 2006). Likewise, long-term effects of smaller classes are also under debate. Many studies have attempted to determine whether there are long-term benefits associated with early childhood interventions such as class-size reductions. Most of these studies report significant benefits that last for a few years after program implementation (Krueger and Whitmore, 2001), but that tend to fade over time (Chetty et al., 2011).

Despite the remarkable amount of literature on class size, almost all studies focus on average outcomes. This focus might appear surprising, especially in light of the undeniable policy relevance of class size effects for children at different levels of achievement. Most studies attempt to deal with heterogeneous effects by using sub-sample analyses (Schanzenbach, 2006). However, evaluating the impact of smaller classes on groups that are only “more likely” to be low achieving (e.g., ethnic minorities or students eligible for free lunches) misses important dimensions of effect heterogeneity (Bitler et al., 2006).

In this chapter, we use data from the Tennessee Student/Teacher Achievement Ratio experiment (also known simply as Project STAR) to examine the distributional effects of being assigned to a small class or a regular class with a teacher’s aide, compared to a regular class. Project STAR—labeled “one of the great experiments in education in U.S. history” by Mosteller et al. (1996)—involved the random assignment of over 11,000 K-3¹ students at roughly 80 public schools to either a small class (i.e., fewer students

¹Pupils from kindergarten through third grade.

than usual), a regular class, or a regular class with a teacher's aide. The experiment began with random assignment of kindergartners and teachers in the mid 1980s, and their assignment to each treatment condition was intended to continue through third grade.

To estimate distributional effects, we employ the unconditional quantile regression approach, recently developed by Firpo et al. (2009). This estimator provides a direct measure of how a marginal change in the level of one variable affects the distribution of achievement in the population, keeping the distribution of other characteristics equal. Unconditional quantile regression differs from conventional quantile regression² in that treatment effects are not conditional on the mean value of included explanatory variables and do not depend on the relative position of an individual among the (virtual) population of individuals who share the same observed characteristics (Fournier and Koske, 2013).

In the second part of the chapter, to estimate the effects of a small class and a regular class with an aide through high school, we use data from Project STAR follow-up surveys. We focus on test score outcomes, on-time high school graduation, and college exam-taking. In particular, we are interested in whether the effects are different for students who were low, mid-, or high achievers during Project STAR, to study, from a distributional point of view, not only the emergence of class size effects but also their persistence.

Our results show considerable heterogeneity in the treatment effects “small class” and “regular class with aide,” suggesting that average effects—while of interest—hide crucial features about the rest of the distribution. We find that mid-achieving pupils profit the most from being assigned to a small class, whereas pupils at the bottom and top of the achievement distribution experience only minimal gains from being in a small class. We also reveal positive and significant effects of an aide in a regular class for low-achieving pupils, i.e., pupils at the bottom 20 percent of the achievement distribution.

²In the sense of Koenker and Bassett (1978).

Interestingly, the effect of having a class aide is strongest for boys and disadvantaged children, to the point that such effect is as strong as that of small classes for the same type of pupils.

Regarding the long-term effects of class size, we analyze three outcomes: test scores in later grades, on-time high school graduation, and college entrance exam-taking. In line with previous research, we find that average effects on test scores in later grades tend to shrink. Adding to the existing literature, we show that the effects in later grades remain strong for Project STAR high achievers, whereas they vanish completely for low and mid-achievers. This result may help resolve the debate on the existence of a long-term effect of class size on test scores, because it suggests that the long-term benefits of being in a small class fade very quickly for low- and mid-achieving students. For the aide treatment, we find no long-term effect on test scores. Concerning the impact of “small class” and “regular class with aide” on on-time high school graduation and college exam-taking, we find positive and significant (average) effects. The effect size of smaller classes and classes with an aide are very similar, further underling the importance of the aide treatment aide.

The remainder of this chapter proceeds as follows. Section 2.2 gives an overview of the theoretical background and empirical literature relevant for class size and Project STAR. Section 2.3 briefly introduces the data set, presents some descriptive statistics, and tests the validity of the experimental design (covariate balance and nonrandom attrition). Section 2.4 shows the econometric models in detail. Section 2.5 presents the results, and section 2.6 concludes.

2.2 Theory and Empirical Background

Education economists have invested a lot of effort in studying class-size effects, mainly for two reasons. First, class size is readily measurable and modifying it constitutes a

relatively easy policy to implement—although expensive. Moreover, both teachers and parents usually perceive class size to be negatively correlated with student achievement. Second, traditional economic theory suggests that smaller classes have a positive impact on achievement (Todd and Wolpin, 2003). In every education production function, class size is always listed along with other relevant school resources such as teacher qualification and school funding (Hanushek, 2002).

In terms of mechanisms, Lazear (2001) argues that disruptive students who take up instructional time in ways not useful to other students affect not only their own learning but also that of classmates. He argues that, in smaller classes, the likelihood of having disruptive students decreases, allowing teachers to spend less time on correcting misbehavior. Furthermore, teachers might find teaching less costly in terms of effort if they have to manage smaller classes, simply because they have fewer students to supervise (Angrist and Lavy, 1999).

Although theoretical research suggests positive effects of smaller classes on student achievement, the empirical evidence is mixed. On the one hand, studies relying on experimental data consistently find positive causal effects of class size reduction on student test scores (Fletcher, 2009; Krueger, 1999, 2003) and non-cognitive skills (Dee and West, 2011; Ding and Lehrer, 2011; DePaola et al., 2013). On the other hand, non-experimental studies present a less optimistic view of class size effects. Indeed, most non-experimental studies report either relatively small class-size effects (Woessmann and West, 2006) or no effects at all (Hoxby, 2000; Woessmann, 2005). The long-term effect of smaller classes is also a matter of debate. While the literature finds positive effects in both elementary and secondary school (Fletcher, 2009; Krueger and Whitmore, 2001), the benefits in the labor market are very small at best (Chetty et al., 2011). Lazear (2001) explains this lack of empirical support in the non-experimental literature by arguing that teachers (and schools) adjust their behavior to smaller classes, therefore bringing no significant effects on test scores.

The effects of class size and pupil-to-teacher-ratio over the achievement distribution are much less analyzed. While we might expect that class size reductions could have different impacts on children with different cognitive abilities, the empirical literature left this question unanswered. Only few recent studies (Jackson and Page, 2013; Konstantopoulos, 2008) have attempted to tackle this heterogeneity issue, but the evidence remains mixed. To better understand the effects of class size reduction on the achievement gap, Konstantopoulos (2008) examines the variance of test scores within smaller and larger classes. He finds that class size reduction increases not only level of achievement but also variance in achievement. In addition, he finds no significant evidence that smaller classes would reduce the achievement gap between low and high achievers.

To estimate the effects over the achievement distribution, Jackson and Page (2013) employ a different econometric approach, finding that the largest test score gains are at the top of the distribution. However, by estimating the class-size effect as a difference between treatment and control groups at each percentile of the observed achievement distribution, they provide only a measure of within-group variation. Their approach thus loses not only the information coming from the rest of the distribution but also restricts the estimates to the conditional distribution of achievement.

In this chapter, we first employ a new approach to estimating the quantile treatment effect of class size over the entire achievement distribution. As we describe in section 2.4, we use the unconditional quantile regression method, which has several advantages over standard quantile regression estimators and other approaches to study distributional effects. Second, as none of the existing studies include a distributional analysis of the treatment “regular size with aide” (pooling such treatment condition with the control group, thus hiding the effect of aide), we also offer a detailed analysis of the effect of “regular size with aide” and its impacts over the achievement distribution. Third, we investigate distributional long-term effects of class size on later grades and high school. Our goal is to investigate the persistence of class size effects and, if so, whether these

long-term effects are heterogeneous for children at different achievement levels.

2.3 Data and Descriptive Statistics

In this section, we briefly overview the data from Project STAR and the follow-up surveys in the later grades and high school. Then we present descriptive statistics and covariate balance. In the last part, to ensure that differential attrition patterns are not endangering our identification strategy, we perform a complete attrition analysis for both Project STAR and later grades.

2.3.1 Project STAR and Follow-up Data

We use data from the Tennessee Student/Teacher Achievement Ratio experiment, a large-scale, randomized class-size experiment that took place between 1985 and 1989 and involved 11,600 students from kindergarten through third grade.³ Project STAR was commissioned by the Tennessee state legislature and implemented by a consortium of Tennessee universities and the Tennessee State Department of Education. The total cost of the experiment, including the cost of hiring new teachers and classroom aides, was approximately USD 12 million.

Initially, all school districts in Tennessee were asked to participate in Project STAR, and about 180 schools in about 50 of the 141 state districts showed an interest. Only about 100 schools had enough students in each grade to meet the size criterion for participation. This size criterion, which was necessary for allowing assignment to class types within schools, excluded very small schools from the study. At the end of this selection process, 79 elementary schools in 42 school districts became part of Project STAR.

Districts had to agree to participate for four years and to allow site visitations for

³Finn and Achilles (1990); Folger and Breda (1989), and Word et al. (1990)—from which we draw heavily in this section—present detailed information about the experiment.

verification of class sizes, interviewing, and data collection. All schools had to allow the random assignment of pupils and teachers to class types from kindergarten through third grade. The experiment randomly assigned kindergarten pupils to small classes (target enrollment between 13 and 17 students), regular classes (target enrollment between 22 and 26 students), or regular classes with a full-time teacher's aide. These teachers' aides did not have to possess any specific educational background and they did not receive any particular training in their duties, their job was helping teachers prepare materials for class and tutoring individual students with learning difficulties. The class-type assignments of students and teachers were maintained through the third grade. Children and teachers entering the study after kindergarten were also randomly assigned to one of the treatments.

Although Project STAR covered only one state and one cohort, the experiment included a heterogeneous set of schools from across Tennessee, including large and small, urban and rural, and wealthy and poor districts. Consequently, the schools included in the STAR data represent most of the educational conditions that exist in the United States (Finn and Achilles, 1990; Krueger and Whitmore, 2001). Additionally, the data includes detailed information on pupils, teachers, and schools, for which we can control for.

The measure of achievement is the Stanford Achievement Test (SAT-9), which all Project STAR children took at the end of each grade. The SAT-9 is a standardized, multiple choice test that includes math, reading, and word identification as subject areas. Because there are no standard units for the test results, we follow Krueger (1999) by scaling the test scores into percentile ranks. Specifically, at each grade level and for each treatment condition, we assign percentile scores based on students' raw test scores, ranging from 0 (lowest score) to 100 (highest score). For each subject test (math, reading, word), we generate a separate percentile distribution and, to summarize overall achievement, we calculate the average of the three SAT-9 percentile rankings for each student.

As a robustness check, we also perform the analysis using standardized scores (z-scores) instead of percentile ranks as in Jackson and Page (2013), with very similar results.⁴

Although after leaving Project STAR all students returned to regular-size classes, they were followed through high school, from which students graduate after successfully completing grades 1–12. Consequently, the students are usually 17-18 years old by the time they finish high school. In their last or next-to-last year of high school, students who intend to enroll in college take the ACT or SAT exam. During grades 4 through 8, the Tennessee Department of Education provided researchers with standardized achievement test scores, and matched such test scores to the appropriate Project STAR cohorts.⁵ Similarly, the College Board, together with the Educational Testing Service, linked information on high school seniors in the class of 1998 who took the ACT or SAT exam to records of the 11,600 children who participated in Project STAR, regardless of where the former Project STAR students resided in 1998. The resulting data set contains high school records of Project STAR students, as well as whether they took either the ACT or SAT exam, and what their test score was.

2.3.2 Covariate Balance in STAR

The main advantage of a randomized experiment is that it provides a solution to the problem of causal inference. In principle, if randomization is done appropriately, the mean outcomes of the treatment and control groups can stand as counterfactuals for one another, making inference about the effects of the treatment relative to the control transparent. If the treatment conditions were truly randomly assigned to pupils in each school, then individuals in the treatment and control groups should have—in expectation—the same pre-intervention characteristics. One way to test this condition is to check whether assignment to a treatment condition is predictive of pupil and teacher characteristics. If

⁴See appendix Table 2.7.3

⁵From fourth grade through eighth, the Tennessee Department of Education provided researchers with standardized achievement tests (the Tennessee Comprehensive Assessment Program).

pupils and teachers were properly randomized within schools, we should observe no statistically significant relationship between each treatment condition and all pre-intervention covariates.

To check for random assignment, which was performed at entry into Project STAR, we disaggregate the data into waves according to the grade in which children entered the experiment.⁶ Then, for each cohort, we regress student and teacher characteristics on the treatment dummies. We also have to include school fixed effects, because random assignment was only valid within schools. We finally perform an F -test of the hypothesis that the class-type dummies had no joint effect.

Table 2.3.1 presents the results, divided according to grade cohorts. Although we focus only on kindergarten and first grade, the results for the second and third grades are very similar.⁷ In Table 2.3.1, the p -values of the class-size dummies exceed 0.05, meaning that there is no significant difference in observable characteristics across treatments. The only exception, of course, is class size itself. In one case the p -value is less than 0.10. It appears that less experienced teachers were assigned to the regular first-grade classes. Nonetheless, the difference is only marginally significant.

Table 2.3.1 also presents attrition rates for each treatment group and for kindergarten and first grade. We define attrition as a binary variable that equals one if a pupil ever left the study, and zero otherwise. Once we have the attrition variable, we regress it on the treatment indicators and the school fixed effects. For pupils who entered Project STAR in first grade, we find no evidence of different attrition rates among treatment groups. However, for pupils who entered in kindergarten, those assigned to a regular class—either with an aide or without—appear more likely to have left the sample.

Because we are also interested in the distributional effects of class size, we check randomization over the achievement distribution. To do so, we divide the achievement distribution into quartiles and regress student and teacher characteristics on the treat-

⁶Krueger (1999) and Krueger and Whitmore (2001) use similar approaches.

⁷Results are available upon request.

Table 2.3.1: DESCRIPTIVE STATISTICS AND COVARIATE BALANCE

	Small (1)	Regular (2)	Aide (3)	Joint p -value (4)
A. Kindergarten				
Class size	15.1	22.4	22.8	0.00**
Girl	0.49	0.49	0.48	0.89
Black	0.31	0.32	0.34	0.45
Age (in 1985)	4.66	4.64	4.65	0.48
Free-lunch eligible	0.47	0.48	0.50	0.51
Black teacher	0.14	0.20	0.15	0.40
Teacher with master	0.31	0.36	0.36	0.65
Teacher experience	9.90	10.1	10.7	0.38
Attrition	0.49	0.52	0.53	0.02*
B. First Grade				
Class size	15.9	22.7	23.5	0.00**
Girl	0.49	0.44	0.46	0.41
Black	0.36	0.43	0.34	0.11
Age (in 1985)	4.98	5.06	5.08	0.21
Free-lunch eligible	0.59	0.62	0.61	0.34
Black teacher	0.18	0.22	0.20	0.58
Teacher with master	0.32	0.33	0.36	0.88
Teacher experience	13.9	11.0	13.2	0.09 [†]
Attrition	0.53	0.51	0.47	0.52

Notes: ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.10$. Ordinary least squares models with robust standard errors clustered at the school level. Sample size in panel A ranges from 5,902 to 6,325 and in panel B ranges from 2,190 to 2,314.

Project STAR data, Authors' calculations.

ment dummies and school fixed effects. Appendix Table 2.7.1 presents the results for the below-median part of the achievement distribution, whereas appendix Table 2.7.2 presents the results for the above-median part of the achievement distribution. As in the mean comparison case, the distributional analysis shows differences in class size across the treatments that are significant at the highest level, suggesting that we are able to identify the effects of class size over the entire distribution. In the first two quartiles of the achievement distribution (appendix table 2.7.1), we observe only a marginally significant under-representation of girls in the regular class group at the bottom quartile in kindergarten ($p = 0.10$), and a marginally significant difference in teacher experience ($p = 0.07$) at the bottom quartile in first grade. For all other pupil and teacher characteristics, the p -values of the class-size dummies are beyond any significance level.

A similar picture emerges from the upper part of the achievement distribution (appendix table 2.7.2), where we observe a 5-percent significant over-representation of girls in the regular class group at the top quartile in kindergarten ($p = 0.02$). The other demographic characteristics have a p -value of the class-size dummies that exceeds 0.05. In first grade we estimate a marginally significant difference in age (third quartile, $p = 0.10$), share of black teachers (third quartile, $p = 0.10$), and teacher experience (top quartile, $p = 0.06$). In the distributional analysis, we find no evidence of different attrition rates across the treatments.

Overall, we find no apparent evidence that initial assignment to class types was highly correlated with either pupil or teacher characteristics. Therefore, we can be reasonably confident that, within schools, both pupils and teachers were randomly assigned to the treatment conditions, and that the randomization holds both at the mean and over the achievement distribution. When we estimate the treatment effects, we nevertheless condition not only on school fixed effects but also on all observable student and teacher characteristics, to increase the precision of the point estimates and control for those few significant differences we find in tables 2.3.1, 2.7.1, and 2.7.2.

2.3.3 Attrition Analysis

In every longitudinal study, one portion of the population of interest leaves the sample. In Project STAR, attrition was relatively high (Hanushek, 1999; Krueger and Whitmore, 2001), and if outcome data are missing for some pupils, we might be concerned that the potential outcomes for those who are observed in the treatment group differ from the potential outcomes for those observed in the control group. For example, if children who were assigned to regular classes and who left the sample had on average higher test scores than children who were assigned to small classes and who also left the sample, then the small class effects will be biased upwards.⁸ Even if attrition is not different across treatment groups, departures could yield analytic samples that vary significantly from the original sample, limiting external validity of estimated causal effects. For example, if girls are more likely to leave the study, then estimates of class size effects on achievement based on a disproportionately male sample may not apply to both sexes.

Although there is no data on students who left Project STAR before the test scores were collected, we can look for evidence of nonrandom attrition by examining differences in observable characteristics across treatment conditions. To do so, we regress an indicator of whether a student ever left STAR on the treatment dummies and an interaction between the treatment dummies and the pre-intervention characteristics. Our definition of attrition includes both children permanently lost to follow-up and others who reappear in the sample after having been missing from one or more intermediate waves after an initial entry. This definition is reasonable because for those pupils who left the study for one or more years, we do not know what treatment they were exposed to during the missing period(s).

Table 2.3.2 presents the p -values of the coefficients on the interaction between the treatment indicator and a given student characteristic. Panel A shows the results for

⁸This differential attrition may occur if high-income families of students in regular classes were more likely to withdraw their children from public schools and send them to private ones, because they would have preferred their children to be assigned to smaller classes.

Table 2.3.2: DIFFERENTIAL ATTRITION IN STAR AND IN FOLLOW-UP SURVEYS

	Girl	Black	Age (in 1985)	Free-lunch Eligible	Percentile Rank Score
	(1)	(2)	(3)	(4)	(5)
A. Assigned to Small					
<i>Project STAR</i>					
Kindergarten	0.20	0.01*	0.38	0.12	0.90
First Grade	0.89	0.20	0.06 [†]	0.19	0.91
Second Grade	0.21	0.39	0.77	0.59	0.29
<i>Later Grades</i>					
Fourth Grade	0.41	0.88	0.05*	0.94	0.56
Eighth Grade	0.22	0.48	0.51	0.47	0.44
High School	0.54	0.50	0.21	0.22	0.17
B. Assigned to Aide					
<i>Project STAR</i>					
Kindergarten	0.10	0.04*	0.24	0.66	0.88
First Grade	0.04*	0.17	0.76	0.20	0.12
Second Grade	0.72	0.83	0.72	0.67	0.99
<i>Later Grades</i>					
Fourth Grade	0.59	0.36	0.07 [†]	0.45	0.96
Eighth Grade	0.53	0.75	0.61	0.26	0.54
High School	0.82	0.40	0.44	0.29	0.16

Notes: ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.10$. Ordinary least squares models with robust standard errors clustered at the school level. Sample size in panel A ranges from 5,902 to 6,833 and in later grades ranges from 10,854 to 10,988.

Project STAR data and follow-up surveys, Authors' calculations.

students originally assigned to small-class treatment, whereas panel B shows the results for students originally assigned to aide treatment. We also distinguish between differential attrition during STAR and differential attrition from follow-up surveys (i.e., fourth grade, eighth grade, and high school).

In general, we find little evidence of nonrandom attrition, at least for the variables we have. Most importantly, we find no evidence of differential attrition in terms of achievement (column 5) and free-lunch eligibility (column 4), suggesting that high-achievers and children from high-income families were not more likely to leave the samples, compared to low-achievers and children from low-income families respectively. This is important because if high-income families (or families of high achieving pupils) were more likely to withdraw their children from school when such children are assigned to a regular class, we would overestimate the effect of the treatments.

For the other variables, some sporadic difference is significant at the five-percent level for girls in the “aide” treatment (in first grade), black students for both “small” and “aide” (in kindergarten), and age in the “small” treatment (in fourth grade). Although these attrition patterns are not present in all the grades, they may cause some problems to the interpretation of our findings if the analytic sample varies significantly from the original—and representative—sample. Given that we find no persistent pattern of nonrandom attrition across all samples, we are confident that this is not the case. Nonetheless, in appendix Table 2.7.3 we adjust for nonrandom attrition by imputing test scores for students who exited the samples. We imply a worst-case scenario, which consists of predicting the scores of pupils who left the control group as if they received the treatment *Small*. Conversely, we predict for pupils who left one of the treatment groups as if they received no treatment. This imputation technique should lead to an increase in average achievement for the control group and, at the same time, a decrease in average achievement for the treatment groups.

2.4 Econometric Models

In this section, we first introduce the education production function that describes the relationship between class size and student achievement. Second, we specify the regression equation we use to estimate the production function of our interest. Third, we describe in detail the econometric method we use to identify the effect of class size across the achievement distribution.

2.4.1 The Education Production Function

To understand the effect of student-to-teacher ratio on achievement and underline the advantages of a randomized experiment for causal inference, we focus on the following education production function (Krueger, 1999; Todd and Wolpin, 2003):

$$Y_{ij} = f(S_{ij}, F_{ij}, R_{ij}) \quad (2.4.1)$$

where Y_{ij} is the achievement level of student i in school j , S_{ij} is a vector of student characteristics, F_{ij} is a vector representing the family background of student i , and R_{ij} is a vector containing school resources and characteristics. If we assume the production function to be linear and separable, we can rewrite equation (2.4.1) as follows:

$$Y_{ij} = a \cdot S_{ij} + b \cdot F_{ij} + c \cdot R_{ij} + \varepsilon_{ij} \quad (2.4.2)$$

where ε_{ij} is the stochastic error component. In principle, S_{ij} , F_{ij} , and R_{ij} include information and experiences that students have been accumulating over their lives. Typical observable characteristics of S_{ij} are, for example, gender and age. In F_{ij} we have variables such as socio-demographic status and family structure. Finally, R_{ij} contains information such as classroom size and teacher qualifications for each year the child attended school.

Additionally, student unobserved ability also contributes to the achievement level

(Ashenfelter and Rouse, 1998). Consequently, in any non-experimental application researchers generally lack data on some relevant school, family, or student characteristics—whether observed or not. These omitted variables will then appear in the error term, and if the omitted variables are correlated with the included variables, then the estimated parameters will be biased. However, if a given characteristic (e.g., class size) is determined by random assignment, it will be independent of the omitted variables. Thus, with random assignment, a simple comparison of achievement between students in small and large classes, or between classes with aide and regular classes, provides an unbiased estimate of the effect of the student-to-teacher ratio on achievement.

We analyze the STAR data by estimating the following regression equation for students in kindergarten and first grade:⁹

$$Y_{ics} = \beta_0 + \beta_1 \cdot Small_{cs} + \beta_2 \cdot Aide_{cs} + X'_{ics}\beta_3 + Z'_{cs}\beta_4 + \alpha_s + \varepsilon_{ics} \quad (2.4.3)$$

where Y_{ics} is the average percentile score on the SAT test of student i in class c at school s , $Small_{cs}$ is a dummy variable indicating whether the student was assigned to a small class that year, $Aide_{cs}$ is a dummy variable indicating whether the student was assigned to a regular-size class with an aide that year, X_{ics} is a vector of observed student covariates (gender, age, ethnicity, and free-lunch eligibility), and Z_{cs} is a vector of observed teacher covariates (gender, ethnicity, years of experience, and qualifications). Given that the randomization was done *within* schools, we also include school fixed effects (α_s). Adding school fixed effects ensures the independence between treatment assignment and other variables.

The coefficients of interest are β_1 and β_2 . They represent the causal effect of being assigned to a small-size or regular-size with aide class on the percentile score of the SAT test.¹⁰ Given that the dependent variable is expressed in percentile ranks, we interpret the

⁹Results for second and third grade show similar patterns (regression outputs are available upon request).

¹⁰In the literature, such effects are referred as “reduced-form” effects or “intention-to-treat” effects

effect size of β_1 and β_2 as a percentage point change in the distribution of achievement.

2.4.2 Estimation of Distributional Effects

Given that we expect the effect of pupil-to-teacher ratio to differ across the achievement distribution, our goal is to go beyond the average effects and study the impact on the overall distribution. In the early 2000s, new methods for estimating counterfactual distributions emerged (Chernozhukov and Hansen, 2013; Fortin et al., 2011; Melly, 2006). The approach that we use, proposed by Firpo et al. (2009), is called unconditional quantile regression.

Firpo, Fortin, and Lemieux’s estimator allows for a direct measure of how a marginal change in the level of one variable (in our case, the treatment dummy) affects the distribution of achievement in the population, keeping the distribution of other characteristics equal. Unconditional quantile regression differs from the commonly used (conditional) quantile regression (Koenker and Bassett, 1978; Koenker and Hallock, 2001). Conditional quantile regression estimates treatment effects conditional on the mean value of included covariates, and the interpretation of such treatment effects change when different sets of covariates are entered into the regression equation. Consequently, the interpretation of effects is limited when the effects for different conditional quantiles vary—e.g., changes in estimated treatment effects may be attributable to either improved identification or estimation of a different relationship. In such cases, the estimated effects do not translate to relevant policy questions that are linked to the covariates of interest (Borah and Basu, 2013). In contrast, unconditional quantile regression can be used for overcoming the limitations of the conditional quantile regression approach.

A simple example adapted from Frölich and Melly (2010) illustrates this advantage. As in our Project STAR setting, assume that the treatment has been completely randomized and is thus independent of both potential outcomes and other covariates. In such a

(Krueger, 1999).

situation, a comparison of the distribution of the outcome in the treated and non-treated populations has a causal interpretation. However, either for efficiency or because of block randomization, we may wish to include covariates or fixed effects in the estimation. If we are interested in mean effects, it is well known that including covariates that are independent of the treatment in a linear regression leaves the estimated treatment effect unchanged. This property is lost for quantile treatment effects: Including covariates that are independent of the treatment changes the limit of the estimated *conditional* quantile treatment effect. However, including those covariates does not change the *unconditional* treatment effect, as long as the exogeneity assumptions of the model are satisfied (which are indeed in our randomized setting).

Additionally, because a conditional quantile is the relative position of an individual among a (virtual) population of individuals that share precisely the same observed characteristics, conditional quantile regression yields only the within-group effect, whereas unconditional quantile regression estimates the total effect, i.e., the sum of the between-group and within-group effects (Fournier and Koske, 2013). This means that unconditional quantile regression allows to compare estimated effects at different quantiles with each other, whereas ordinary quantile regression does not allow for such comparison.

The unconditional quantile regression consists of running a regression of a relatively simple transformation—the re-centered influence function—of the outcome variable on the explanatory variables. Because of its policy relevant interpretation and its computational attractiveness, unconditional quantile regression has been used in several studies on quantile effects (Maclean et al., 2014; Müller, 2015; Stueber and Beissinger, 2012), decomposition analyses (Heywood and Parent, 2012; Sakellariou, 2012; Tang and Long, 2013), and regression-discontinuity designs (Frandsen, 2012).

To understand the unconditional quantile regression method, let Y denote the outcome of interest (e.g., test score) and $F_Y(y)$ denote the cumulative distribution function of Y in a target population. Empirical researchers are often interested in exploring the effect

of a binary treatment D on Y , and denote the potential outcomes and their distribution function under alternate values of D as Y_0 and Y_1 . Thus we have $Y = (1 - D) \cdot Y_0 + D \cdot Y_1$. As Borah and Basu (2013) show, we can express the unconditional distribution function for Y as a weighted average of conditional distribution of Y given D , weighted by the unconditional distribution of D :¹¹

$$F_Y(y) = Pr(D = 1) \cdot F_{Y|D}(y|D = 1) + Pr(D = 0) \cdot F_{Y|D}(y|D = 0) \quad (2.4.4)$$

In most regression models that study the relationship between Y and D , the focus is on the conditional expectation of Y , i.e., $E(Y|D) = \int dF(Y|D)$. In the traditional linear model, ordinary least squares (OLS) is a consistent estimator of the target parameter representing the effect of D on Y , which is simply the difference in the conditional expectation of Y for $D = 0$ and $D = 1$: $\beta_{ols} = E(Y|D = 1) - E(Y|D = 0)$. For OLS, β_{ols} is also a consistent estimator for this marginal effect on the unconditional distribution of Y :

$$\begin{aligned} E(Y) &= p(D) \cdot E(Y|D = 1) + [1 - p(D)] \cdot E(Y|D = 0) \\ dE(Y)/dp(D) &= E(Y|D = 1) - E(Y|D = 0) = \beta_{ols} \end{aligned} \quad (2.4.5)$$

This duality of interpretation of β_{ols} persists—under regularity conditions—even when we add other covariates to the model.

As with OLS, the definition of the effect on the unconditional quantile does not change with the set of covariates available for conditioning: Even in the presence of a vector of covariates, the effect on the unconditional quantile is always evaluated marginally over the distribution of such covariates. Without any other covariates correlated with the outcome, the conditional and the unconditional treatment effects of a treatment variable D are identical for any quantile of Y (Frölich and Melly, 2010). However, when the data-

¹¹In statistics, the *unconditional* distribution of a random variable is referred to as the *marginal* distribution of that variable. However, because we use the term “marginal” to represent small changes in covariate values (marginal effects), we use the term “unconditional distribution” as in Firpo et al. (2009).

generating process allows for the effects to vary over the values of other covariates, the definition of the unconditional quantile treatment effect—as well as the interpretation of such effects—is different from the definition of the conditional quantile treatment effect, because the latter varies with the set of covariates included (Maclean et al., 2014).

For clarity, we follow Borah and Basu (2013) and consider the relationship of D with the quantiles of the distribution of Y . Letting $q_Y(\tau)$ denote the τ^{th} quantile of the unconditional distribution of Y ¹² and following equation (2.4.4) we have that:

$$F_Y[q_Y(\tau)] = Pr(D = 1) \cdot F_{Y|D=1}[q_Y(\tau)] + Pr(D = 0) \cdot F_{Y|D=0}[q_Y(\tau)] \quad (2.4.6)$$

The unconditional quantile treatment effect is obtained by the differentiation of (2.4.6):¹³

$$dq_Y(\tau)/dp(D) = [F_{Y|D=1}(q_Y(\tau)) - F_{Y|D=0}(q_Y(\tau))]/f_Y(q_Y(\tau)) \quad (2.4.7)$$

From Koenker and Hallock (2001), we know that the coefficient on a binary variable D from a conditional quantile regression that includes a vector of covariates X is:

$$\beta_{cqr}(\tau) = F_{Y|D=1, X=\bar{x}}(\tau) - F_{Y|D=0, X=\bar{x}}(\tau) \quad (2.4.8)$$

where \bar{x} represents a vector of sample means for X . $\beta_{cqr}(\tau)$ is a consistent estimator for the conditional effect of D evaluated at the mean values of X , but it is not a consistent estimator of the unconditional effect of D as defined in equation (2.4.7). This result is due to the fact that $dq_Y(\tau)/dp(D) \neq \beta_{cqr}(\tau)$.¹⁴ For example, the 90th percentile of the unconditional distribution of Y may not be the same as the 90th percentile on the conditional distribution of $Y|D, X$.

To provide practitioners with a way to compute $dq_Y(\tau)/dp(D)$, the literature has developed several approaches, from non-parametric estimators (Firpo, 2007) to propensity

¹² $\tau = F_Y[q_Y(\tau)]$

¹³See Firpo et al. (2009).

¹⁴Alternatively, $F_{Y|D=1, X=\bar{x}}(\tau) = q_{Y|D=1, X=\bar{x}}(\tau) \neq q_Y(\tau)$.

score matching (Frölich, 2007) and decomposition methods (Machado and Mata, 2005). The most recent approach—the one we use in this chapter—was proposed by Firpo et al. (2009), who suggest a unconditional quantile regression model based on the concept of re-centered influence function, commonly used in robust statistics (Hampel et al., 2011).

An influence function is an analytic tool assessing the effect of removing or adding an observation on the value of a certain statistic $v(F)$, without having to recalculate that statistic. The influence function is defined as follows:

$$IF[y, v(F)] = \lim_{h \rightarrow 0} \frac{v[(1-h) \cdot F + h \cdot \delta_y] - v(F)}{h}, 0 \leq h \leq 1 \quad (2.4.9)$$

where F represent the cumulative distribution function for Y , and δ_y is a distribution that puts mass at the value y . We obtain the re-centered influence function (RIF) by adding the statistics of interest to its influence function:

$$RIF(y, v) = v(F) + IF(y, v) \quad (2.4.10)$$

If the statistic of interest is a specific quantile τ of the distribution of the outcome of interest, we have:

$$IF[y, v(F)] = (\tau - I[Y \leq q_\tau]) / f_Y(q_\tau) \quad (2.4.11)$$

where q_τ is the τ^{th} quantile of the unconditional distribution of Y , $f_Y(q_\tau)$ is the probability density function of Y evaluated at q_τ , and $I[Y \leq q_\tau]$ indicates whether an outcome value is less than the specified quantile q_τ . In the case of quantiles, the re-centered influence function is then:

$$RIF(y, q_\tau) = q_\tau + IF(y, q_\tau) \quad (2.4.12)$$

Firpo et al. (2009) show that when the conditional expectation of $RIF(y, q_\tau)$ is modeled as a function of explanatory variables, a RIF regression can be viewed as an unconditional quantile regression.¹⁵

¹⁵This is because, as $E_X E[RIF(Y, \tau) | x] = q_\tau$ by the definition of RIF, Firpo et al. (2009) demonstrate

In the class size experiment context, relying on unconditional quantile regressions—rather than ordinary quantile regressions—provides researchers and policymakers with additional information. For example, we might be interested in studying the impact of teacher or student characteristics on achievement and how class size effects are moderated by such variables. This is only possible with unconditional quantile regression methods, because including covariates leaves the treatment effect unchanged. This property also exists in ordinary least squares models, but it is lost for conventional quantile regression methods. Another example concerns the interpretation of the estimated coefficients. The class size effect estimated by conventional quantile regression methods only shows the within-group variation of the effects. This means that the coefficient indicates, for each quantile, how much variation there is in the class size effect among a group with the same student and teacher characteristics. From a policy perspective this is not informative, because we would be more interested in the total effect, i.e., the within and between variation. This is only possible with unconditional quantile regression methods. Therefore, unconditional quantile regression is not only more suitable to answer policy questions but it also allows comparing estimated effects at different quintiles with each other.

The implementation of the unconditional quantile regression, is a computationally attractive two-step procedure.¹⁶ For a specific quantile τ , we first have to estimate the RIF of the τ^{th} quantile of Y following (2.4.11) and (2.4.12). We calculate q_τ using the sample estimate of the unconditional τ^{th} quantile of Y . Similarly, we estimate the density $f_Y(q_\tau)$ at q_τ using kernel methods. The second step is to run an OLS regression of the $RIF(y, q_\tau)$ on the treatment variables and other observed covariates. In this two-step procedure, the unconditional quantile partial effects are simply the estimated coefficients.¹⁷

that $E_X[dm_\tau(x)/dX]$ is the marginal effect of a covariate on the τ^{th} unconditional quantile of Y , *ceteris paribus*.

¹⁶See Firpo et al. (2009) for the detailed procedure and alternative approaches.

¹⁷To compute the unconditional quantile treatment effects, we use the Stata routine `rifreg`, available at <http://faculty.arts.ubc.ca/nfortin/datahead.html>

2.5 Results

This section presents our results in two parts. The first part shows the results for student test scores using the unconditional quantile regression. Given that the distributional investigation constitutes the core of this chapter, we present detailed regression outputs and sub-sample analysis, both supported by graphical evidence. The second part examines class-size effects in later grades. We focus on fourth grade, eighth grade, and high school.¹⁸

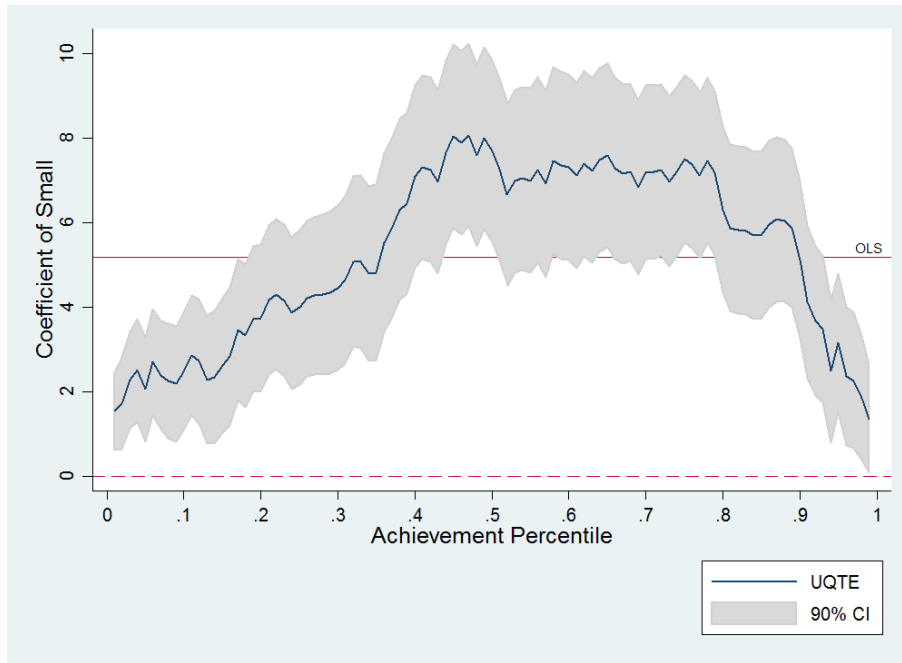
2.5.1 Distributional Effects During Project STAR

Table 2.5.1 presents the unconditional treatment effects of smaller classes (i.e., *Small*) and classes with a teacher's aide (i.e., *Aide*) on achievement in kindergarten and first grade. We gradually add pupil and teacher covariates to our regression equations, and although doing so does not largely alter the effect size, pupil and teacher characteristics are jointly significant. Thus our preferred estimates are those in column 3 for kindergarten and in column 6 for first grade.

A look at the coefficients of *Small* in Table 2.5.1 reveals that being assigned to a small class has positive effects on test scores throughout the entire achievement distribution ($p = 0.00$). However, there is a high degree of heterogeneity in the effects. At the bottom decile of the distribution, students assigned to smaller classes score 2.5 percentage points higher than those assigned to regular-size classes in kindergarten. In first grade, bottom-decile students assigned to smaller classes score 4.5 percentage points higher than those assigned to regular-size classes. The effect is larger at the median, reaching 7.7 percentage points in kindergarten and 8.5 percentage points in first grade. Then the effect of *Small* declines in the upper part of the achievement distribution, down to 5-6 percentage points. Then the effect of *Small* declines in the upper part of the achievement distribution, down

¹⁸Results for fifth, sixth, and seventh grade—which are very similar to those of fourth and eighth—are available upon request.

Figure 2.5.1: Distributional Effect of Small Class, Kindergarten



to 5-6 percentage points.

In Figures 2.5.1 and Figure 2.5.2, we compute the unconditional quantile treatment effect of *Small* for each percentile of the achievement distribution, along with the respective 90-percent confidence intervals. Both figures further highlight the heterogeneous effect of being assigned to a small class. It shows that the mean effect—the continuous red line—is a poor representation of the small-class effect. Mid-achievers (fourth to eighth decile) profit the most from being assigned to a small class. Pupils at the bottom and at the top of the achievement distribution experience only a little from being in a smaller class.

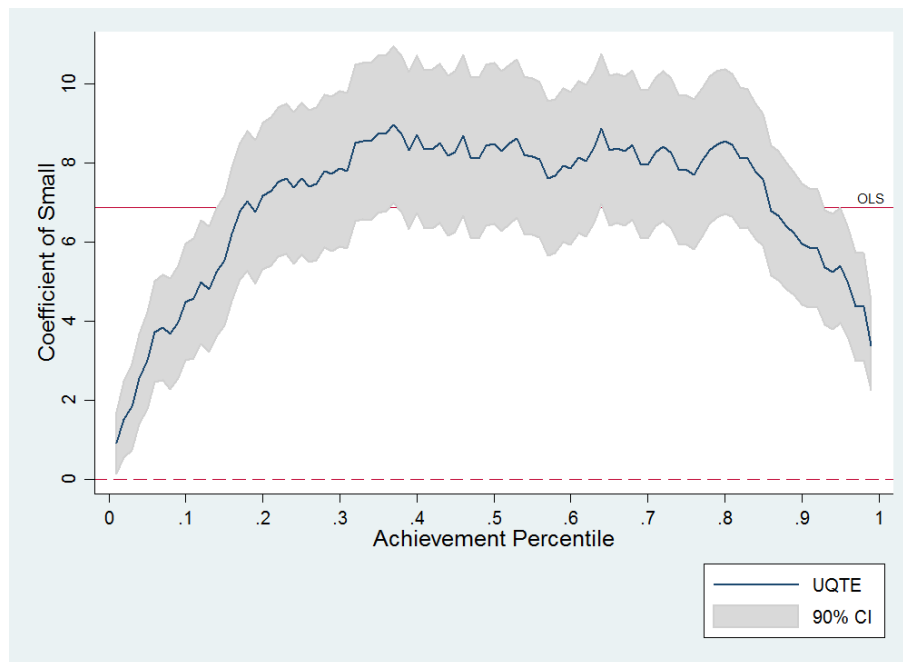
Table 2.5.1 also presents unconditional quantile treatment effects for the treatment condition *Aide*. Consistent with previous studies, we find no effect on test score at the median. However, low achievers actually benefit from aide-teachers. For pupils at the bottom of the achievement distribution, we estimate a positive and highly significant effect of *Aide*. The effect size is roughly 2.5 percentage points on the percentile rank score, an effect as large as that of being in a small class in kindergarten. To emphasize the importance of *Aide*, figures 2.5.3 and 2.5.4 present unconditional quantile treatment

Table 2.5.1: CLASS SIZE EFFECTS IN KINDERGARTEN AND FIRST GRADE

Percentile Rank Score						
	Kindergarten			First Grade		
	(1)	(2)	(3)	(4)	(5)	(6)
Quantile .10						
Small class	2.423** (0.856)	2.349** (0.849)	2.476** (0.843)	5.179** (0.899)	4.679** (0.891)	4.476** (0.898)
Aide class	2.266** (0.832)	2.346** (0.823)	2.462** (0.824)	2.532** (0.913)	2.595** (0.904)	2.323** (0.900)
<i>Adjusted R²</i>	<i>0.096</i>	<i>0.116</i>	<i>0.117</i>	<i>0.082</i>	<i>0.099</i>	<i>0.101</i>
Quantile .25						
Small class	4.083** (1.132)	3.899** (1.108)	3.992** (1.109)	9.372** (1.168)	8.358** (1.145)	7.603** (1.165)
Aide class	1.326 (1.100)	1.485 (1.073)	1.353 (1.076)	2.581* (1.171)	2.637* (1.147)	1.743 (1.159)
<i>Adjusted R²</i>	<i>0.159</i>	<i>0.197</i>	<i>0.199</i>	<i>0.143</i>	<i>0.179</i>	<i>0.182</i>
Quantile .50						
Small class	7.770** (1.348)	7.536** (1.300)	7.700** (1.307)	10.528** (1.255)	9.171** (1.218)	8.487** (1.238)
Aide class	-1.070 (1.291)	-0.845 (1.249)	-0.864 (1.263)	2.502* (1.206)	2.552* (1.172)	1.823 (1.192)
<i>Adjusted R²</i>	<i>0.163</i>	<i>0.217</i>	<i>0.218</i>	<i>0.164</i>	<i>0.215</i>	<i>0.216</i>
Quantile .75						
Small class	7.760** (1.224)	7.521** (1.193)	7.510** (1.203)	9.217** (1.170)	8.230** (1.143)	7.820** (1.151)
Aide class	-0.007 (1.124)	0.145 (1.095)	-0.153 (1.107)	1.484 (1.091)	1.519 (1.061)	1.101 (1.083)
<i>Adjusted R²</i>	<i>0.143</i>	<i>0.184</i>	<i>0.185</i>	<i>0.132</i>	<i>0.176</i>	<i>0.178</i>
Quantile .90						
Small class	5.167** (1.136)	5.034** (1.118)	5.121** (1.126)	6.456** (0.941)	5.986** (0.929)	5.949** (0.931)
Aide class	-0.967 (0.973)	-0.867 (0.957)	-0.949 (0.973)	0.303 (0.809)	0.316 (0.796)	0.306 (0.817)
<i>Adjusted R²</i>	<i>0.084</i>	<i>0.110</i>	<i>0.110</i>	<i>0.081</i>	<i>0.107</i>	<i>0.107</i>
School fixed effects	YES	YES	YES	YES	YES	YES
Student covariates		YES	YES		YES	YES
Teacher covariates			YES			YES
<i>N</i>	5,837	5,837	5,837	6,449	6,449	6,449

Notes: ** $p < 0.01$, * $p < 0.05$, $^{\dagger} p < 0.10$. Robust standard errors are in parentheses. Student covariates include gender, age, age squared, ethnicity, and free-lunch eligibility. Teacher covariates include gender, ethnicity, years of experience, experience squared, and qualifications. Project STAR data, Authors' calculations.

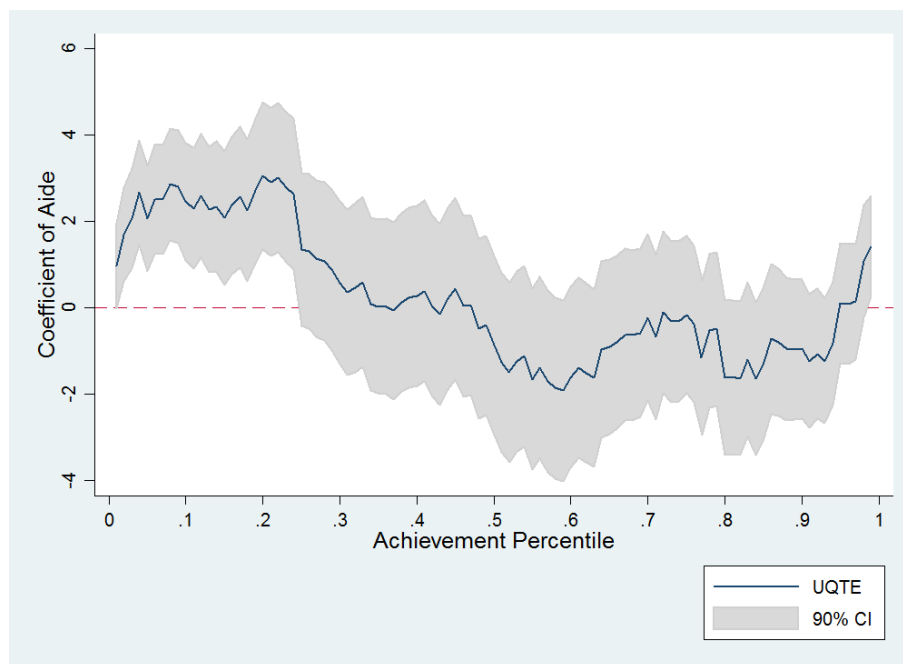
Figure 2.5.2: Distributional Effect of Small Class, First Grade



effect at each percentile of the achievement distribution.

In kindergarten (figure 2.5.3), being assigned to a regular class with an aide clearly has a positive effect on test score for low-achieving students. The effect is highly significant for the first two deciles of the achievement distribution and ranges between 2-3 percentage points. For the rest of the distribution, the treatment condition *Aide* has no impact on test scores. In first grade (figure 2.5.4), the effect of *Aide* is irregular and imprecisely estimated. The reason has to do with the experimental design (Krueger, 1999). After one year of Project STAR (i.e., kindergarten), a first assessment of the treatment *Aide* revealed no significant effect on test scores (at the mean). Therefore, the experimental committee decided to re-randomize pupils who were originally assigned to one of the treatments in a regular-size class. This re-randomization makes it impossible for researchers to know precisely which treatment the pupils in regular classes experienced. Therefore, most studies of Project STAR do not even analyze the effects of being assigned to *Aide*; instead, they usually pool *Aide* with the control group. Despite this experimental flaw, we can investigate the impact of *Aide* at least in kindergarten, allowing us to reveal positive effects for low achievers. The benefits of *Aide* are present in first grade,

Figure 2.5.3: Distributional Effect of Aide, Kindergarten



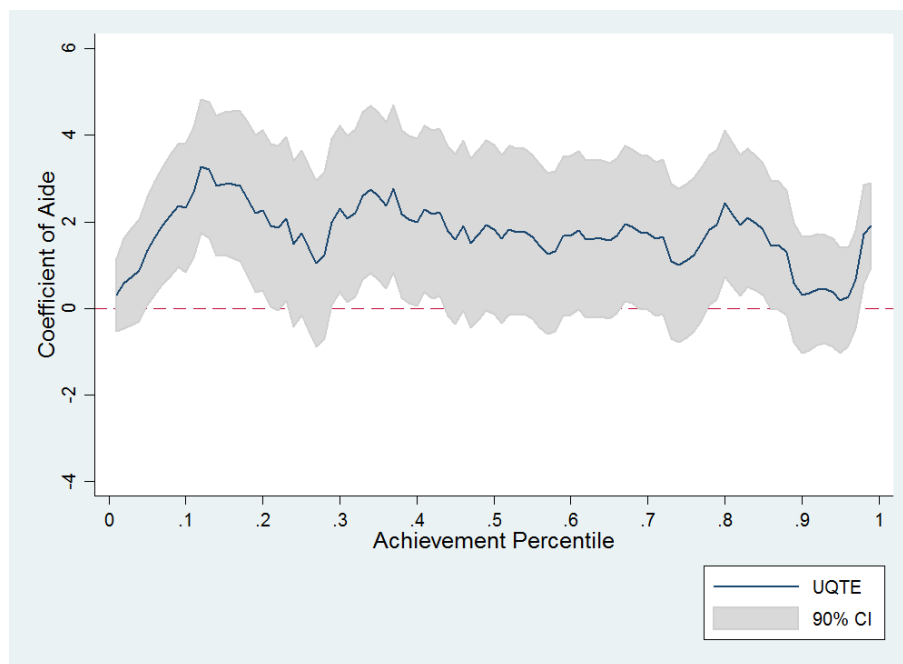
too, but they are less clear and less precisely estimated.

To further assess whether the effects of *Small* and *Aide* are beneficial for disadvantaged pupils, we estimate unconditional quantile treatment effects for black and for free-lunch eligible children. We also conduct a separate analysis for boys, because the literature from education economics and psychology shows that boys tend to be more disruptive (Bertrand and Pan, 2013) and are more likely to lose their concentration during instructional time (Feingold, 1994).¹⁹

Table 2.5.2 presents the sub-sample analysis for both kindergarten and first grade outcomes. A smaller class has larger effects for boys, black children, and free-lunch eligible children. This is consistent with previous studies that focus on average effects, but our analysis also underlines a high level of heterogeneity in the effect of *Small*. The inverted u-shaped pattern over the percentiles of the achievement distribution persists, but now the effect drops only at the top decile (instead of the top quintile). Similarly, being assigned to a regular class with a teacher's aide is very beneficial for boys and disadvantaged

¹⁹We also performed the analysis for white pupils and girls. We do not report the results here, because they are very similar to the estimates using the full sample (slightly smaller effect sizes). However, results are available upon request.

Figure 2.5.4: Distributional Effect of Aide, First Grade



children. The treatment *Aide* has a strong positive impact on low-achievers' test scores and such effect is significant almost for half of the achievement distribution.

Figures 2.5.5 provides graphical evidence of the effect of *Small* and *Aide* for boys, black children, and free-lunch eligible children. We restrict the analysis to kindergarten not only for the sake of comprehension but also to focus on the outcomes that were measured before the re-randomization. Comparing the effect of *Aide* for the sub-samples to the one of the entire sample, we observe two phenomena. First, the effect is larger in terms of magnitude; and, second, the effect is significant up to the fourth decile (whereas for the full sample it is significant only for the bottom two deciles).

We might suspect that the distributional results we obtain depend on how we specified the outcome variable, i.e., in percentile ranks. However, specifying the dependent variable in z-scores (standard deviations) does not affect the results, as we show in appendix Table 2.7.3. As a further robustness check, in appendix Table 2.7.3 we adjust for nonrandom attrition by imputing test scores for students who left the sample. We predict the scores of pupils who left the control group as if they received the treatment *Small*. Conversely, we predict for pupils who left one of the treatment groups as if they received

Table 2.5.2: CLASS SIZE EFFECTS IN KINDERGARTEN AND FIRST GRADE, SUB-SAMPLE

Percentile Rank Score						
	Kindergarten			First Grade		
	Black (1)	Free-lunch (2)	Boys (3)	Black (4)	Free-lunch (5)	Boys (6)
Quantile .10						
Small class	3.680** (1.249)	3.267** (1.001)	2.093 [†] (1.136)	4.335** (1.031)	3.327** (0.917)	3.975** (1.182)
Aide class	3.543** (1.192)	3.191** (0.977)	4.045** (1.081)	1.353 (1.166)	1.406 (0.929)	2.463* (1.143)
<i>Adjusted R</i> ²	0.086	0.072	0.125	0.060	0.043	0.083
Quantile .25						
Small class	6.552** (1.561)	5.563** (1.274)	5.042** (1.414)	7.454** (1.410)	6.673** (1.268)	7.401** (1.652)
Aide class	3.387* (1.495)	3.798** (1.259)	4.930** (1.356)	2.620 [†] (1.511)	3.331** (1.252)	3.145* (1.601)
<i>Adjusted R</i> ²	0.168	0.136	0.191	0.134	0.103	0.166
Quantile .50						
Small class	8.377** (2.473)	7.050** (1.903)	9.700** (1.820)	10.518** (1.866)	8.461** (1.631)	6.650** (1.711)
Aide class	-2.290 (2.252)	-0.657 (1.799)	3.090 [†] (1.754)	-1.083 (1.858)	1.429 (1.514)	1.551 (1.620)
<i>Adjusted R</i> ²	0.218	0.182	0.211	0.167	0.150	0.207
Quantile .75						
Small class	8.824** (2.693)	7.850** (2.083)	9.772** (1.830)	11.656** (2.454)	7.352** (1.831)	8.504** (1.679)
Aide class	-1.549 (2.301)	-1.111 (1.884)	0.559 (1.682)	-1.211 (2.210)	1.865 (1.668)	0.728 (1.580)
<i>Adjusted R</i> ²	0.199	0.180	0.182	0.151	0.139	0.187
Quantile .90						
Small class	7.568** (2.583)	7.094** (2.085)	9.525** (1.716)	9.757** (2.470)	5.927** (1.968)	5.718** (1.406)
Aide class	1.676 (2.177)	0.816 (1.770)	1.531 (1.450)	-2.516 (2.009)	0.514 (1.739)	0.051 (1.215)
<i>Adjusted R</i> ²	0.156	0.126	0.128	0.097	0.094	0.114
School fixed effects	YES	YES	YES	YES	YES	YES
Student covariates	YES	YES	YES	YES	YES	YES
Teacher covariates	YES	YES	YES	YES	YES	YES
<i>N</i>	1,897	2,823	2,991	2,143	3,289	3,338

Notes: ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.10$. Robust standard errors are in parentheses. Student covariates include gender, age, age squared, ethnicity, and free-lunch eligibility. Teacher covariates include gender, ethnicity, years of experience, experience squared, and qualifications.

Project STAR data, Authors' calculations.

Figure 2.5.5: Distributional Effect of Small and Aide on Test Scores, Sub-Samples



no treatment. Doing so reduces the estimated treatment effects—as expected—but does not entirely wipe out either the small-class effect or the aide-class effect.

Given the heterogeneity we observe in the data for both *Small* and *Aide*, we need to explain the mechanism behind our results. Understanding why high achievers benefit less from a small class compared to the mid-achievers is somehow intuitive. Being a high achiever usually correlates with both higher motivation and higher socioeconomic status (Heckman and Masterov, 2007). Children with such characteristics would probably perform well regardless of the treatment they receive, or at least they would benefit less compared to those students who lack of either resources or motivation.

Understanding why low achievers benefit less from being in smaller classes, instead, might be less intuitive. Theory would suggest that in small classes teachers are more able to identify low achievers and thus more likely to provide instruction designed to benefit these students (Konstantopoulos, 2008). However, such practice is difficult to implement when there is only one teacher in the classroom, because he or she would need to focus only on a group of children, leaving the majority of the class without supervision. Therefore, teachers in smaller classes probably use their additional capacity to help the majority to improve.

By contrast, when the teachers are two as in the *Aide* treatment, teachers are not only likely to identify low achievers but also able to provide targeted instruction to benefit such students. This might explain the positive impact of being in a class with an extra teacher for those pupils who need more help or support. This result has an important policy implication: Not only the treatment *Aide* actually has an impact on low-achievers' test scores but such effect is also stronger for boys and disadvantaged children. Therefore, for low-achieving classes it might be beneficial to hire an extra professional to provide in-class aide.

An alternative explanation for the minimal effect of *Small* among low achievers might come from empirical observation. An unfortunate recurrence in many interventions

that aim at low achievers is that such interventions obtain only scarce or partial success (Betts and Shkolnik, 2000; Jacob and Lefgren, 2009; Lefgren, 2004), probably because low achievers and their families are less responsive to treatments in general. However, the fact that low achievers benefit most from the treatment *Aide* suggests that this latter explanation cannot be the entire story, at least in our case.

2.5.2 Effects in Later Grades and High School

In later grades we cannot perform the same distributional analysis as in the STAR years, because assuming rank preservation in the achievement distribution between kindergarten and fourth grade or eighth grade is likely too stringent. We nevertheless want to investigate whether class-size effects are different for low, mid-, and high achievers. To do so, we add an indicator of whether a student was a bottom, middle, or top achiever during the STAR years, and we then interact the indicator with the treatment dummies. Unfortunately, in later grades we do not have information about teacher characteristics, partly because students have different teachers for several subjects (especially in high school).

Table 2.5.3 presents the class size effects on fourth and eighth grade. In columns 1 and 3 we see that being in a smaller class during Project STAR has a positive effect in later grades. The effect size is 2.7 percentage points in fourth grade, and declines until 2.2 at the end of eighth grade. Krueger and Whitmore (2001) and Schanzenbach (2006) also report similar decreasing effects, whereas Nye et al. (1999) estimate no fading between third grade and eighth grade. The difference in the results is probably due to different econometric approaches, but none of the studies investigate whether the effects differ among low, mid-, or high achievers.

Columns 2 and 4 show the regression outputs for the models that include interaction terms between treatment conditions and achievement dummies. Three findings are relevant in this analysis. First, being a mid- or high achiever during Project STAR has a huge impact on later grades test scores, compared to low achievers. The effect size is

Table 2.5.3: CLASS SIZE EFFECTS IN FOURTH AND EIGHTH GRADES

	Percentile Rank Score			
	Fourth Grade		Eighth Grade	
	(1)	(2)	(3)	(4)
Small class	2.737*	-3.867**	2.170*	-3.140*
	(1.174)	(1.142)	(1.036)	(1.328)
Aide class	-0.058	-1.376	0.142	-1.150
	(1.001)	(1.071)	(0.927)	(1.275)
Mid-achiever		23.763**		17.348**
		(1.593)		(1.600)
High achiever		45.973**		36.165**
		(1.73)		(1.405)
Small*Mid-achiever		-0.454		-0.717
		(2.022)		(1.852)
Small*High achiever		3.967*		3.887*
		(1.929)		(1.798)
Aide*Mid-achiever		-2.054		-0.836
		(1.758)		(2.060)
Aide*High achiever		2.133		2.000
		(1.652)		(1.570)
School fixed effects	YES	YES	YES	YES
Student covariates	YES	YES	YES	YES
(Adjusted) R ²	0.280	0.645	0.308	0.543
N	4,043	4,043	5,056	5,056

Notes: ** $p < 0.01$, * $p < 0.05$, † $p < 0.10$. Ordinary least squares models with robust standard errors clustered at the school level are in parentheses. Student covariates include gender, age, age squared, ethnicity, and free-lunch eligibility.

Project STAR data and follow-up surveys, Authors' calculations.

around 20 percentage points for mid-achievers and it doubles for high achievers. Second, the only significant coefficient among all interactions is the one on *Small*High achiever* ($p < 0.05$), suggesting that all the benefits of being assigned to a smaller class in later grades comes from high-achieving students. High achievers benefited less from smaller classes *during* STAR compared to mid-achievers, however high achievers experience much less fading in later grades. Third, we find no significant effect for either *Aide* or any interaction between *Aide* and the achievement dummies. This might seem disappointing, but it might be due to the fact that we have accurate information about the *Aide* treatment only for kindergarten. Moreover, in later years, the confidence bounds of the interaction terms are large, suggesting that we might lack of statistical power to achieve significance.

We conduct a similar analysis for high school, using on-time graduation and ACT/SAT exam-taking as outcomes. Both outcomes are coded as binary variables, and Table 2.5.4 summarizes the results. We find that both *Small* and *Aide* have a positive and marginally significant effect on high school graduation ($p < 0.10$), and effect of 2.9 percentage points for *Small* and 2.3 percentage points for *Aide*. The effects are not moderated by any of the achievement dummies, as we see in column 2. However, we observe that including the interactions we lose precision in estimating the coefficients. This might be due either to a lack of statistical power or to a high correlation between the treatment variables and the achievement dummies.

Regarding ACT/SAT exam-taking, we confirm the findings of Krueger and Whitmore (2001), i.e., that being assigned to smaller classes in early grades increases the likelihood of taking a college entrance exam. Moreover, we additionally find a positive and significant effect of *Aide*, which constitutes a relevant extension to Krueger's work. Both coefficients are highly significant ($p < 0.01$), with a magnitude of roughly 5 percentage points. The interaction between *Small* and *High achiever* is marginally significant ($p < 0.10$), which might imply that—once again—the effect of *Small* is more persistent for high-achieving students, compared to low and mid-achievers. Compared to test scores results, it appears

Table 2.5.4: CLASS SIZE EFFECTS IN HIGH SCHOOL

	Graduated from High School		Took ACT/SAT in High School	
	Dummy=1 if Yes		Dummy=1 if Yes	
	(1)	(2)	(3)	(4)
Small class	0.029 [†] (0.015)	0.046 (0.032)	0.051** (0.013)	-0.012 (0.016)
Aide class	0.023 [†] (0.013)	0.025 (0.023)	0.044** (0.011)	0.029* (0.013)
Mid-achiever		0.115** (0.033)		0.174** (0.021)
High achiever		0.142** (0.028)		0.373** (0.019)
Small*Mid-achiever		-0.071 (0.046)		0.030 (0.030)
Small*High achiever		-0.026 (0.044)		0.052 [†] (0.027)
Aide*Mid-achiever		-0.023 (0.037)		-0.004 (0.021)
Aide*High achiever		-0.009 (0.033)		0.011 (0.023)
School fixed effects	YES	YES	YES	YES
Student covariates	YES	YES	YES	YES
(Adjusted) R ²	0.158	0.171	0.156	0.247
N	4,667	4,667	10,335	10,335

Notes: ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.10$. Ordinary least squares models with robust standard errors clustered at the school level are in parentheses. Student covariates include gender, age, age squared, ethnicity, and free-lunch eligibility.

Project STAR data and follow-up surveys, Authors' calculations.

that *Aide* helps complete high school on time but does not necessarily improve test scores. This is consistent with previous findings on college entrance exam-taking behavior, which indicates that class size increases the probability of taking an ACT or SAT exam but not the result of the test itself Krueger and Whitmore (2001).

2.6 Conclusions and Discussion

Using data from Tennessee’s Project STAR and follow-up surveys, we provide experimental evidence of heterogeneous effects of class size and teacher’s aide over the achievement distribution. Our results contribute to the literature on class size in at least four ways. First, we show that, given the large amount of heterogeneity in the treatment effects, mean regression provides only a poor description of the underlying relationship between class size and achievement. Similarly, not even standard sub-sample analysis is a sufficient tool for studying heterogeneity and heterogeneity patterns over the achievement distribution.

Second, we find that mid-achieving students gain the most from being assigned to a small class, whereas students at the bottom and top of the achievement distribution experience only minimal gains. Although this result differs from what those few studies that investigate the distributional effects of class size suggest, we use a richer econometric approach (i.e., unconditional quantile regression), and our findings are robust across alternative specifications and estimation techniques.

Third, we report—for the first time—positive and significant effects of a regular-size class with an aide for the low-achieving pupils. Not only is the effect significant for the first two deciles of the achievement distribution, but it is even stronger for boys and disadvantaged children. Interestingly, the effect size of *Aide* is as large as that of *Small* for the bottom third of the achievement distribution. In terms of equity, while the net effect of *Small* on the achievement gap is not clear, our estimates show that adding a

teacher's aide would be an effective policy for reducing the achievement gap, especially for classes with large percentages of boys, black students, or low-income students.

Fourth, our results confirm that average effects on test scores in later grades, on-time high school graduation, and college exam-taking are relatively weak and short-lived. However, we are able to show in addition that the effects in later grades remain strong for Project STAR high achievers, whereas they vanish completely for low- and mid-achieving pupils. This result helps resolving the (empirical) debate on the existence of long-term effect of class size, because it clearly shows that the effect depends on the individual's position in the achievement distribution. For high achievers, even though the positive effect of smaller classes is initially lower than that for mid-achievers in early grades, these benefits persist over time.

Our analyses have some limitations.²⁰ We acknowledge two potential methodological improvements and one potential conceptual advancement. Methodologically, we estimate a reduced-form effect, not the treatment effect on the treated. One way of obtaining that treatment effect would be instrumenting the class size that an individual actually experienced with his or her initial assignment. However, the unconditional quantile regression estimator of Firpo et al. (2009) applies only to models with no endogenous regressors. Another issue related to the econometric approach we use are the standard errors. The optimal strategy, given the data structure, would be computing robust standard errors clustered at the school level. However, as our approach does not allow clustering, we compute robust standard errors. Future research could attempt to apply block-bootstrapping methods (Cameron et al., 2008).

Conceptually, several researchers argue that test scores are not the most relevant outcome in education production functions. For example Heckman et al. (2013) stress

²⁰We skip the discussion on data limitations, as it is already been discussed extensively (Hanushek, 1998, 1999). The main limitation of the Project STAR data is that baseline test scores are not available. Thus testing that the treatment and control groups were similar in terms of their achievement distributions before the experiment began is not possible. However, the random assignment should ensure equivalence before the intervention. In our analyses we find no worrisome observable differences between the groups.

the importance of non-cognitive skills, which have a significant impact on labor market outcomes. Similarly, studies by Duckworth et al.,²¹ emphasize the role of perseverance and passion for long-term goal on education and labor market success. Given that follow-up surveys of Project STAR contain enough information to build a “grit score,” future research might explore this avenue in more detail.

This chapter shows that typical estimates of the mean gain from class size and teacher’s aide provide an incomplete characterization of their real impact on achievement distribution, thus constituting a weak guide for public educational policy. While smaller classes have the largest impact on mid-achievers, having an in-class teacher’s aide constitutes an effective measure for raising the test scores of low achievers. Similarly, while a teacher’s aide appears to have no impact on test scores at the mean, having a teacher’s aide is extremely useful for low-achieving pupils. We conclude that policymakers, when designing educational reforms, need to think carefully about the goals and the impact that an intervention has on different parts of the achievement distribution, rather than its impact on the mean achiever.

²¹See for example Duckworth et al. (2007) and Eskreis-Winkler et al. (2014).

2.7 Appendix

2.7.1 Descriptive Statistics over the Achievement Distribution

Table 2.7.1: DESCRIPTIVE STATISTICS AND COVARIATE BALANCE OVER THE DISTRIBUTION (BELOW THE MEDIAN)

	Bottom Quartile				Second Quartile			
	Small (1)	Regular (2)	Aide (3)	Joint p -value (4)	Small (5)	Regular (6)	Aide (7)	Joint p -value (8)
A. Kindergarten								
Class size	15.2	22.6	22.7	0.00**	15.1	22.4	22.7	0.00**
Girl	0.44	0.40	0.45	0.10 [†]	0.49	0.45	0.46	0.59
Black	0.42	0.47	0.49	0.90	0.34	0.29	0.35	0.83
Age (in 1985)	4.64	4.61	4.62	0.46	4.64	4.64	4.65	0.78
Free-lunch eligible	0.65	0.68	0.71	0.74	0.55	0.49	0.54	0.33
Black teacher	0.15	0.25	0.13	0.11	0.14	0.17	0.15	0.66
Teacher with master	0.27	0.38	0.33	0.36	0.31	0.41	0.35	0.58
Teacher experience	9.17	8.89	9.60	0.36	9.95	10.4	10.4	0.78
Attrition	0.74	0.73	0.70	0.53	0.54	0.51	0.55	0.59
B. First Grade								
Class size	15.6	22.8	23.7	0.00**	15.6	22.4	23.3	0.00**
Girl	0.50	0.42	0.49	0.37	0.42	0.46	0.42	0.93
Black	0.57	0.61	0.51	0.45	0.45	0.50	0.44	0.61
Age (in 1985)	5.07	5.11	5.12	0.92	5.14	5.10	5.14	0.99
Free-lunch eligible	0.79	0.81	0.78	0.66	0.68	0.68	0.67	0.37
Black teacher	0.33	0.26	0.32	0.52	0.19	0.22	0.23	0.89
Teacher with master	0.28	0.26	0.37	0.45	0.26	0.37	0.34	0.39
Teacher experience	12.4	9.90	13.4	0.07 [†]	13.5	11.9	13.4	0.68
Attrition	0.78	0.65	0.63	0.41	0.55	0.53	0.47	0.52

Notes: ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.10$. Ordinary least squares models with robust standard errors clustered at the school level.

Sample size in panel A ranges from 5,902 to 6,325 and in panel B from 2,190 to 2,314.

Project STAR data, Authors' calculations.

Table 2.7.2: DESCRIPTIVE STATISTICS AND COVARIATE BALANCE OVER THE DISTRIBUTION (ABOVE THE MEDIAN)

	Third Quartile				Top Quartile			
	Small (1)	Regular (2)	Aide (3)	Joint p -value (4)	Small (5)	Regular (6)	Aide (7)	Joint p -value (8)
A. Kindergarten								
Class size	15.1	22.2	22.8	0.00**	15.1	22.3	22.7	0.00**
Girl	0.50	0.53	0.49	0.26	0.51	0.61	0.55	0.02*
Black	0.30	0.29	0.26	0.77	0.22	0.21	0.22	0.12
Age (in 1985)	4.65	4.64	4.63	0.76	4.68	4.67	4.69	0.80
Free-lunch eligible	0.45	0.41	0.40	0.71	0.31	0.29	0.31	0.63
Black teacher	0.17	0.18	0.16	0.92	0.10	0.16	0.14	0.78
Teacher with master	0.33	0.32	0.39	0.67	0.34	0.34	0.40	0.89
Teacher experience	10.2	10.2	11.0	0.69	10.1	11.0	12.3	0.16
Attrition	0.39	0.41	0.43	0.34	0.33	0.36	0.37	0.20
B. First Grade								
Class size	16.0	22.5	23.5	0.00**	16.0	22.9	23.1	0.00**
Girl	0.48	0.41	0.49	0.35	0.52	0.49	0.46	0.43
Black	0.29	0.30	0.30	0.93	0.23	0.26	0.18	0.27
Age (in 1985)	4.87	5.03	5.07	0.10 [†]	4.85	4.96	4.98	0.34
Free-lunch eligible	0.51	0.55	0.55	0.29	0.41	0.38	0.44	0.83
Black teacher	0.12	0.20	0.15	0.10 [†]	0.13	0.18	0.12	0.93
Teacher with master	0.34	0.34	0.35	0.91	0.40	0.38	0.38	0.71
Teacher experience	15.0	11.8	13.6	0.26	14.8	10.8	12.2	0.06 [†]
Attrition	0.43	0.38	0.39	0.60	0.33	0.39	0.33	0.39

Notes: ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.10$. Ordinary least squares models with robust standard errors clustered at the school level.

Sample size in panel A ranges from 5,902 to 6,325 and in panel B from 2,190 to 2,314.

Project STAR data, Authors' calculations.

2.7.2 Robustness Checks

Table 2.7.3: ROBUSTNESS CHECKS, Z-SCORES AND IMPUTED SCORES

	Z-Scores (standard deviations)		Imputed Test Scores	
	Kindergarten (1)	First grade (2)	Kindergarten (3)	First grade (4)
<i>Quantile .10</i>				
Small class	0.110** (0.037)	0.285** (0.063)	2.604** (0.858)	2.693** (1.019)
Aide class	0.125** (0.036)	0.130* (0.062)	2.337** (0.843)	2.007* (0.969)
<i>Quantile .25</i>				
Small class	0.157** (0.037)	0.307** (0.050)	3.439** (1.068)	6.193** (1.286)
Aide class	0.063 [†] (0.036)	0.113* (0.049)	0.696 (1.048)	2.143 [†] (1.252)
<i>Quantile .50</i>				
Small class	0.251** (0.041)	0.247** (0.041)	5.853** (1.147)	7.767** (1.337)
Aide class	-0.014 (0.039)	0.012 (0.040)	-1.437 (1.108)	2.441 [†] (1.269)
<i>Quantile .75</i>				
Small class	0.312** (0.047)	0.166** (0.029)	6.917** (1.184)	6.798** (1.200)
Aide class	-0.025 (0.043)	0.019 (0.027)	-0.048 (1.095)	1.111 (1.123)
<i>Quantile .90</i>				
Small class	0.236** (0.064)	0.166** (0.025)	5.581** (1.141)	5.593** (0.976)
Aide class	-0.086 (0.055)	0.008 (0.021)	-0.735 (0.985)	0.534 (0.854)
School fixed effects	YES	YES	YES	YES
Student covariates	YES	YES	YES	YES
Teacher covariates	YES	YES	YES	YES
<i>N</i>	5,837	6,449	6,253	6,455

Notes: ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.10$. Robust standard errors are in parentheses. Student covariates include gender, age, age squared, ethnicity, and free-lunch eligibility. Teacher covariates include gender, ethnicity, years of experience, experience squared, and qualifications. Project STAR data, Authors' calculations.

Chapter 3

Returns to Education over the Wage Distribution

[A version of this paper has been submitted to *Labour Economics*]

3.1 Introduction

Although a positive relationship between education and wages is one of the standard results in economic literature (Dickson and Harmon, 2011), the question of whether education affects individuals differently over the wage distribution is much less analyzed (Wang, 2013). Moreover, in such distributional settings, the literature has not investigated whether different types of education—vocational or academic—result in differing returns, or whether one type of education brings a return premium compared to the other at some point of the wage distribution but not at others. These questions are particularly important because a lack of information about educational types may lead to costly

decisions for both the individual and the government.

To fill these gaps, in this chapter we first causally estimate the returns to education over the wage distribution. The analysis reveals potential heterogeneous effects of education on wages, answering the question of whether the returns are increasing, decreasing, or u-shaped across the quantiles.

In a second step, we compare the returns to one extra year of academic education with the returns to one extra year of vocational education, to investigate whether one track brings a return premium at any point in the wage distribution. Such a comparison is lacking in the literature, generally because most countries do not have an extensive vocational education and training system that allows acquiring the same quality of education and the same number of years as in the academic track, or because the academic track is more prestigious or preferred than the vocational one.¹

One notable exception is Switzerland,² a country with an extensive vocational and training system that attracts two-thirds of the individuals in every cohort (Tuor and Backes-Gellner, 2010). The Swiss educational system allows students to achieve tertiary education degrees for both academic and vocational tracks. Therefore, using Swiss data allows us to shed light on heterogeneous returns to different types of education keeping constant the level, and to answer the question of how academic and vocational education differ over the wage distribution.

The analyses that we propose address two major issues that are common for estimations of returns to education: endogeneity of education attainment (Harmon et al., 2003) and heterogeneity in the returns to education (Henderson et al., 2011). While theoretical research considers both issues simultaneously (Arias et al., 2001; Card, 1999), empirical work often deals with only one issue at a time. To overcome the endogeneity problem, most scholars use instrumental variable estimation (Angrist and Krueger, 1991; Dickson, 2013; Harmon and Walker, 2000; Trostel et al., 2002).

¹See, e.g., Bettinger et al. (2010) for Colombia.

²Other countries with similar vocational systems are Denmark and Germany (Hanushek et al., 2011).

However, when dealing with the heterogeneity issue, the literature has not converged to a standard method for integrating it into the analysis (Lemieux, 2008). Therefore, researchers usually rely on different methods when accounting for heterogeneity in returns to education: Sub-sample analysis (Harmon et al., 2003), non-parametric estimation (Henderson et al., 2011), Bayesian hierarchical models (Koop and Tobias, 2004), and quantile regression (Fasih et al., 2012; Martins and Pereira, 2004). The first three methods focus mainly on the existence and the nature of heterogeneity, which are not the focus of this chapter. However, quantile regression (QR) is instead more appropriate to our research question, because QR estimates the returns to education over the wage distribution, allowing for heterogeneity through quantile-specific intercepts and quantile-specific slopes.

The use of QR in returns to education studies was hindered for many years because the endogeneity problem in QR models could not be solved. However, recent studies by Chernozhukov and Hansen (Chernozhukov and Hansen, 2008, 2013) propose an instrumental variable quantile regression (IVQR) approach that deals with both heterogeneity and endogeneity at the same time. Although the IVQR method has been applied in many research fields in economics (Atella et al., 2008; Autor et al., 2012; Eren, 2009; Lamarche, 2011; Maynard and Qiu, 2009; Wehby et al., 2009), it is relatively new to the returns to education literature. Only two studies implement IVQR to propose alternative instruments for schooling (Arabsheibani and Staneva, 2012) and to examine the inequality-reducing effect of education in China (Wang, 2013).

Exploiting a major education reform that took place in Switzerland in the 1970s, we use IVQR to causally estimate the returns to education over the wage distribution, and we compare the results with standard QR and ordinary least squares (OLS) to determine whether taking endogeneity into account changes results and conclusions. In a second step, we also distinguish between educational paths, to add a new comparison between and within academic and vocational education. In this latter comparison we are especially

interested in the presence of heterogeneity, and we therefore use only conventional QR methods.³

The results provide evidence that there is no unique causal effect of schooling and that for each individual the effect may deviate from those extensively documented by ordinary least squares or two-stage least squares. In particular, while ordinary quantile regression estimates increasing returns in the quantile index, once the endogeneity of schooling is taken into account we instead observe higher returns at lower quantiles of the wage distribution.

We also reveal significant heterogeneity within the academic and the vocational track, and comparing these two paths shows that academic education does not always yield higher returns. In the upper half of the wage distribution, individuals with an academic background have higher returns than individuals with a vocational background. However, at lower quantiles of the wage distribution, vocational education brings higher returns than academic education, suggesting that answering the question of which type of education has larger returns is not as easy as it might appear from descriptive statistics or mean regression.

The remainder of this chapter proceeds as follows. Section 3.2 gives an overview of the theoretical background related to our research questions. Section 3.3 introduces the data set and presents descriptive statistics. Section 3.4 shows the econometric models in detail. Section 3.5 presents the results, and section 3.6 concludes.

3.2 Theory and Empirical Background

In this section, we briefly present some theoretical background and empirical evidence to derive our hypotheses and provide a structure for our empirical analysis. We follow the theoretical model developed by Card (1999); its most interesting feature is that it

³Nevertheless, we also performed instrumental variable (quantile) regressions, which are presented in appendix Table 3.7.6.

considers both heterogeneity in the returns and endogeneity of education attainment in the wage equation at the same time.

Following Card, we assume that individuals choose their level of education to maximize the following utility function defined over wage and years of education:

$$U(w, S) = \ln(w) - f(S) = \ln[g(S)] - f(S) \quad (3.2.1)$$

where $g(S)$ and $f(S)$ are increasing convex functions that represent the benefits and costs of schooling, respectively. The condition $w = g(S)$ captures the observable relationship of wage to schooling, i.e., the level of wages available at each level of education. The first order condition for optimal education is:

$$\frac{g'(S)}{g(S)} = f'(S) \quad (3.2.2)$$

In the optimum, the marginal rate of return to education equals the marginal cost. Individual heterogeneity in the optimal education choice arises from two sources: differences in the cost of education, represented by the variation in $f(S)$, and differences in the monetary benefit of education, represented by the variation in $g'(S)/g(S)$.

To characterize the well-documented fact that (log)wage is a nearly linear function of schooling that may vary across individuals,⁴ we impose the following functional form to the heterogeneity components:

$$MB_i = \frac{g'(S)}{g(S)} = b_i - k_1 \cdot S_i \quad (3.2.3)$$

$$MC_i = f'(S) = r_i + k_2 \cdot S_i \quad (3.2.4)$$

where b_i and r_i are random variables with some joint distribution across the population

⁴Card and Krueger (1992), Heckman and Polachek (1974), and Hungerford and Solon (1987) present evidence suggesting that wages are nearly log-linear with respect to schooling. Furthermore, Park (1994) finds log-linearity to be a good approximation of the wage-schooling relationship not only at the mean but also for several quantiles of the wage distribution.

$i = 1, 2, \dots$ and k_1 and k_2 are non-negative constants. To derive an equation for the natural logarithm of wage, we need to integrate the expression for the marginal rate of return to education with respect to S_i :

$$\ln(w_i) = a_i + b_i \cdot S_i - \frac{1}{2} \cdot k_1 \cdot S_i^2 \quad (3.2.5)$$

where a_i is an individual-specific constant of integration.

Equation (3.2.5) is a general version of the functional form adopted in Mincer (1974). However, the salient feature of Card's model is that individual heterogeneity potentially affects both the intercept of the wage equation (through a_i) and the slope of the wage-education relation (through b_i).

This latter feature introduces three important issues into the empirical work. First, we should expect different returns to education for individuals with different levels of ability. More specifically, given that individuals acquire education up to the point where the marginal cost equals the marginal rate of return, and given that costs depend negatively on ability, we should observe that returns to education decrease as ability increases. This means that, while higher-ability individuals have on average higher wages, the slope of their wage-education profile is flatter than that for lower-ability individuals. Second, we cannot assess the true impact of education on wages without solving the bias introduced by the endogeneity of schooling attainment, because otherwise cross-sectional estimates are (marginally) upward biased by an omitted ability variable (Heckman et al., 2006). Third, if we want to study how education affects different individuals, we need to account simultaneously for heterogeneity and endogeneity.

To incorporate these features into our analysis, we use IVQR, which estimates the causal effect of education on conditional quantiles of the wage distribution, allowing for quantile-specific intercepts and quantile-specific slopes. Given that IVQR is a relatively new method, the vast majority of the literature uses conventional QR to investigate the heterogeneous effects of education on wage (Fasih et al., 2012; Harmon et al., 2003;

Hartog et al., 2001; Martins and Pereira, 2004). From these studies we conclude that returns to education vary substantially over the wage distribution. These studies also suggest that returns to education increase in the quantiles of wage distribution. As we can interpret the quantile index as a measure of ability (Arias et al., 2001; Mwabu and Schultz, 1996), this finding contrasts with what we would theoretically expect. However, the implicit assumption of exogenous schooling in conventional QR studies may explain the discrepancy between theoretical expectation and empirical findings.

The few studies applying IVQR in the return-to-education context present mixed results. Using spouse education as an instrument for education, Wang (2013) investigates the evolution of the returns in China over time, to examine the inequality-reducing effect of education. He estimates slightly decreasing returns to education over the wage distribution, ranging from 5.1 percent at the lowest quartile to 3.1 percent at the highest quartile. Proposing risky sexual behavior at an early age as a new instrument for schooling, Arabsheibani and Staneva (2012) apply IVQR to Russian data and find increasing returns over the wage distribution. Specifically, they estimate a 5 percent return at the lowest decile and a 15 percent return at the highest decile. However, when estimating the causal return to education both approaches rely on a demand-side variation in schooling, making defending the orthogonality between the instruments and the error term of the wage equation very difficult. In our study, by contrast, we look into a supply shock as source of exogenous variation in years of education, which is a much more reliable tool to identify the causal effect of education on wages (Arcand et al., 2005).

Pushing the analysis of heterogeneous returns one step further, researchers and policymakers are often interested in the return to different educational types, such as academic and vocational education. While most studies on returns to education do not consider the content of the variable years of education, policymakers—as well as students and parents—may need more information than simply the average return to one year of education. In this context, a typical question is whether vocational education yields a

lower or higher labor market return than an academic education of the same number of years. Furthermore, such type-specific return might well be different for individuals at different points of the wage (ability) distribution.

So far, the few existing papers seem to suggest that academic degrees have generally larger benefits than vocational degrees. Dearden et al. (2002) provide evidence on the relative value of academic and vocational qualifications in the British labor market. Their results appear to show that the wage premium associated with academic qualifications is on average higher than that associated with vocational qualifications at the same level. Similarly, Saniter (2012) examines the returns to education for different educational groups in Germany. He comes to the conclusion that the return to education is 8.5 percent for the entire sample, 2.3 percent for graduates from the basic school track (vocationally oriented), and 11 percent for graduates from a higher school track (academically oriented). Focusing on non-monetary benefits of educational tracks, Hanushek et al. (2011) conclude that gains in youth employment from vocational education are offset by less adaptability and consequent diminished employment later in life. Thus, over the life-cycle, academic education seems to have larger non-monetary benefits than vocational education.

However, none of these studies analyze the return to one extra year of academic education with the return to one extra year of vocational education to investigate whether one type brings a return premium over the other type, nor do they explore the possibility of heterogeneous effects between and within educational paths. This is probably the case because many countries do not have an education system that allows acquiring all levels of education, including tertiary degrees in either the academic or the vocational track. However, Switzerland has both academic and vocational education at all levels, so with our study we are able to add novel results to the discussion on academic versus vocational track by revealing the heterogeneous effects of the two educational types.

3.3 Data and Descriptive Statistics

Before presenting the data and providing descriptive statistics, we briefly introduce the current Swiss education system, which is needed to understand how we can address the endogeneity and heterogeneity issues. The education system in Switzerland consists of parallel paths divided into vocational and academic education. After nine years of compulsory schooling, about two-thirds of a youth cohort choose to pursue vocational education and training, mostly within what is called the “dual system” of apprenticeship training. This kind of training generally comprises a curriculum-based on-the-job training component and a theoretical component taught at specialized vocational schools. After graduation, most of these apprentices work as skilled workers within their occupational fields. Vocational graduates also have several options for continuing their education. They may choose to go into higher vocational education and acquire a higher vocational education degree or a university of applied sciences degree (Tuor and Backes-Gellner, 2010).

On the other hand, about one fifth of a youth cohort chooses to pursue an academic track. They remain in the academic school system, attend academic secondary school and obtain a “Matura,” a high school diploma that is a prerequisite for tertiary academic education. Afterwards, they can acquire degrees at tertiary academic institutions such as universities and federal institutes of technology and earn a bachelor’s degree, a master’s degree, or a doctorate. Therefore, in the Swiss system we find students with, for example, nine years of education in a vocational track along with students with nine years of pure academic track. Switzerland thus has the best setting to answer the research questions we raised in the previous section.

We base our analysis on data from the Swiss Labor Force Survey (SLFS), produced annually by the Swiss Federal Statistical Office. The data are collected by telephone interviews, and the sample is representative for the adult population permanently living in Switzerland. The main purpose of the SLFS is to provide information on employ-

ment patterns and on the structure of the labor force. Strict adherence to international definitions makes Swiss data comparable with OECD, European, and U.S. data. The SLFS was conducted for the first time in 1991 and is based on a sample of about 105,000 interviews. We select the period 2000-2009, and we pool these cross-sections to build our sample.⁵ We take the 2000-2009 period because the SLFS renewed its questionnaire in 1999 and because since 2010 the SLFS has been issued quarterly.

To avoid special circumstances such as those that might arise from retirement, our sample takes into account only males aged 18-60. We also restrict the sample to employed individuals to avoid misspecification resulting from people being in school or not being active in the labor force. Among the employed, to retain individuals with attachment to the labor market, we focus on fully employed workers.⁶ The wage variable of the SLFS comes from the Swiss Survey on Income and Living Conditions, a very precise data source for income resulting from labor activity. We also exclude 0.5 percent of each tail end of the wage distribution to attenuate the impact of outliers and remove implausible values. Wages are expressed in Swiss Francs (CHF) throughout the entire chapter, inflated to the year 2010.⁷

In the SLFS, for each individual, we can observe the entire educational path from compulsory education to doctorate, and we dichotomize the educational paths into academic and vocational according to the official definition of the Swiss State Secretariat for Education, Research and Innovation (appendix figure 3.7.1). After removing individuals with missing values, we are left with 34,744 observations in the sample. Tables 3.3.1 and 3.3.2 provide descriptive statistics.⁸

From the descriptive analysis on the full sample (table 3.3.1), we observe that, over the 2000-2009 period, the average worker earned an annual wage of CHF 81,868 and has

⁵The SLFS is a rotating panel. We keep one observation per individual to prevent problems of nonrandom attrition and clustering.

⁶We use the official definition of the Swiss Federal Statistical Office, which considers an individual as fully employed if he or she has an employment of at least 90 percent.

⁷In 2010, 1 CHF = 1 USD. In Switzerland, inflation is very low and stable over time.

⁸See appendix Table 3.7.1 for the details on sample construction.

Table 3.3.1: DESCRIPTIVE STATISTICS

Variables	Mean (1)	Std. Dev. (2)	Min (3)	Max (4)
Annual wage	81,868	41,824	12,816	390,292
Age	40.23	10.09	18.00	60.00
Years of education	13.16	2.88	7.00	21.00
Years of vocational	2.82	2.34	0.00	9.00
Years of academic	1.80	3.56	0.00	12.50
Compulsory education	0.12	0.33	0.00	1.00
Vocational education	0.65	0.48	0.00	1.00
Academic education	0.23	0.42	0.00	1.00
<i>N</i>	34,744			

Notes: Swiss Labor Force Survey, Authors' calculations.

acquired 13.16 years of education. In line with the statistics at the national level, in our sample 65 percent of the individuals followed a vocational path, whereas 23 percent obtained an academic degree (Tuor and Backes-Gellner, 2010). The rest of the sample (12 percent) has compulsory education as the highest educational level. Table 3.3.2 presents descriptive statistics over the wage distribution, which shows the well-known positive relationship between education and wage. However, these figures do not take into account unobserved heterogeneity; in particular, differences in ability are not factored in. Therefore, descriptive results give no indication of the causal wage effects of different types of education.

3.4 Methods

In this section, we first introduce the equations to be estimated. We use two different models: one to analyze the return to education and one to compare the academic track with the vocational track. Second, we briefly describe the estimation methods we apply, i.e., OLS, QR, and instrumental variable estimations. Given that QR and IVQR are not as common as OLS and two-stage least squares (TSLS), we give a brief overview of these two methods following Koenker and Bassett (1978) and Chernozhukov and Hansen

Table 3.3.2: DESCRIPTIVE STATISTICS OVER WAGE DISTRIBUTION

Variables	Bottom Quartile		Second Quartile		Third Quartile		Top Quartile	
	Mean (1)	Std. Dev. (2)	Mean (3)	Std. Dev. (4)	Mean (5)	Std. Dev. (6)	Mean (7)	Std. Dev. (8)
Annual wage	46,391	8,484	62,830	4,026	80,892	7,065	137,432	46,327
Age	36.79	11.05	39.14	9.94	40.93	9.22	44.06	8.53
Years of education	11.76	2.48	12.11	2.31	13.49	2.50	15.28	2.77
Years of vocational	2.31	2.04	2.87	1.97	3.47	2.27	2.66	2.84
Years of academic	0.80	2.30	0.71	2.25	1.54	3.36	4.16	4.61
Compulsory education	0.27	0.45	0.16	0.37	0.04	0.20	0.01	0.10
Vocational education	0.60	0.49	0.73	0.45	0.77	0.42	0.51	0.50
Academic education	0.13	0.34	0.11	0.31	0.19	0.39	0.48	0.50
<i>N</i>	8,719		8,656		8,684		8,685	

Notes: Swiss Labor Force Survey, Authors' calculations.

(2013). Third, we describe and discuss the instrumental variables we use for the causal estimation of the returns to education.

3.4.1 The Wage Equations

To estimate the private monetary return to one additional year of education, we consider the following Mincer-like equation:

$$\ln(w_i) = \delta_0 + \beta_S \cdot S_i + \delta_1 \cdot Age_i + \delta_2 \cdot Age_i^2 + \varphi_t + u_i \quad (3.4.1)$$

In equation (3.4.1), w_i is the annual wage of individual i , S_i represents the years of education, Age_i is a proxy for labor market experience, φ_t is a set of time controls, and u_i is an error term. As is common in the literature, we exclude various determinants of wages such as tenure and industry sector, because such variables are potentially endogenous and determined by education itself (Angrist and Pischke, 2008). In model (3.4.1), the coefficient of interest is the one on the variable years of schooling β_S , which we expect to be positive and significant.

To compare the effect of one additional year of academic education to the effect of one additional year of vocational education, we develop a model similar to that used by Hartog et al. (2001) and Vandebussche et al. (2006). Hartog et al. modify the classical Mincer wage equation and include a spline in year of education for three categories of the school system: primary, secondary, and tertiary education. With this specification, they investigate the different effects of education on wages among different levels of education. With a similar specification, Vandebussche et al. study the effect of tertiary education on the growth rate of countries. They separate the effect of tertiary education from primary and secondary education to show that skilled labor has a higher growth-enhancing effect for countries closer to the technological frontier. In our case, we decompose the education variable as defined in model (3.4.1) into its three components: compulsory education (C),

vocational education (V), and academic education (A). Thus, we can rewrite equation (3.4.1) as follows:

$$\ln(w_i) = \delta_0 + \beta_C \cdot C_i + \beta_V \cdot V_i + \beta_A \cdot A_i + \delta_1 \cdot Age_i + \delta_2 \cdot Age_i^2 + \varphi_t + u_i \quad (3.4.2)$$

In model (3.4.2), the parameters of interest are β_V and β_A . With this second specification, we compare the return premium of one additional year of vocational education with the premium of one additional year of academic education.⁹ While expecting both parameters to be significant and positive is reasonable, building expectations about the comparison between the two is not straightforward, for the following two reasons. First, previous literature on the topic is scarce. Existing studies either compare higher tracks with lower tracks (Saniter, 2012) or focus on non-monetary returns (Hanushek et al., 2011). Second, the returns to the vocational and academic education may be heterogeneous over the wage distribution, making predictions on the comparison between them difficult to formulate.

3.4.2 Instrumental Variable Quantile Regression

The vast majority of applied econometrics focuses on averages, and such focus partly reflects the difficulty of producing credible average causal effects. As long as the dependent variable is binary, the mean describes the entire distribution. However, many variables such as earnings have continuous distributions, which can change in response to treatments in ways that averages do not fully reveal. QR provides a straightforward, powerful tool for modeling distributional effects, even if the underlying mechanism is complex and multidimensional (Angrist and Pischke, 2008).

To allow for heterogeneous effects of education on wages, we consider the τ^{th} condi-

⁹To test whether the two coefficients are different, we perform an F -test, whose null hypothesis is: $\hat{\beta}_V - \hat{\beta}_A = 0$.

tional quantile wage function hereafter (we drop the indexes for clarity):

$$Q_{\ln(w)}[\tau|X, S] = X'\alpha(\tau) + \beta(\tau)S \quad (3.4.3)$$

where X denotes all explanatory variables other than education $(1, Age_i, Age_i^2, \varphi_t)$, $\alpha(\tau)$ is the return to X at the τ^{th} quantile, $\beta(\tau)$ is the return to education at the τ^{th} quantile, and $\tau \in (0, 1) \mapsto X'\alpha(\tau) + \beta(\tau)S$ is strictly increasing in τ . In equation (3.4.3) the returns to education are a function of τ , allowing for heterogeneous effects of education on wages.

Assuming the error term in the wage equation to be independent of X and S , Koenker and Bassett (1978) propose finding the best predictor of log-wage given X and S under the asymmetric least absolute deviation loss. Doing so means estimating $\alpha(\tau)$ and $\beta(\tau)$ in equation (3.4.3) by solving the following minimization problem:

$$Q_{\ln(w)}[\tau|X, S] = \arg \min_{\alpha(\tau), \beta(\tau)} E[\rho_\tau(\ln(w) - X'\alpha(\tau) - \beta(\tau)S)] \quad (3.4.4)$$

where $\rho_\tau(u_i)$ is the “check function” defined as $\rho_\tau(u_i) = [\tau - \mathbf{1}(u_i \leq 0)]u_i$. In practice, the minimization problem is solved via linear programming and implemented in many statistical packages. As previously discussed, assuming independence between the education variable and the error term may be too stringent because of potential unobserved wage determinants (i.e., ability). To account for potential dependence between S and u in a distributional framework, we apply the IVQR method developed by Chernozhukov and Hansen (Chernozhukov and Hansen, 2006, 2008, 2013).

As in the case of TSLS, the identification of the IVQR approach relies on the existence of a vector Z of instrumental variables that are statistically related to S but independent of the error term u . Additionally, we have to assume that, given the information (X, S) , the distribution of the structural error does not vary across the endogenous state S (“rank

similarity”).¹⁰ The structural error is responsible for heterogeneity of potential outcomes among individuals with the same observed characteristics. This error term determines the relative ranking of observationally equivalent individuals in the distribution of potential outcomes conditional on the individual’s observed characteristics. Rank similarity differs from exact rank invariance by allowing deviations in the individual rank away from some common level. In such formulation, we assume that an individual selects an education level without knowing the exact potential outcomes. Unfortunately, we cannot test rank similarity, but this assumption is consistent with many empirical situations where the exact latent outcomes are not known before a certain treatment (Heckman and Smith, 1997).

Chernozhukov and Hansen show that assuming rank similarity implies the following moment condition:

$$\mathbb{P}[\ln(w) \leq Q_{\ln(w)}(\tau|X, S)|X, Z] = \tau \quad (3.4.5)$$

and thus, in our case:

$$\mathbb{P}[\ln(w) - X'\alpha(\tau) - \beta(\tau)S \leq 0|X, Z] = \tau \quad (3.4.6)$$

The moment condition given in (3.4.6) provides a statistical restriction for use in estimating the parameters $\alpha(\tau)$ and $\beta(\tau)$. Pointing out that equation (3.4.6) is equivalent to the statement that zero is the τ^{th} quantile of the random variable $\ln(w) - Q_{\ln(w)}(\tau|X, S)$ conditional on (X, Z) , Chernozhukov and Hansen formulate the problem as finding $[\alpha(\tau), \beta(\tau)]$ so that zero is the solution to the standard quantile regression of $[\ln(w) - X'\alpha(\tau) - \beta(\tau)S]$ on (X, Z) :

$$0 = \arg \min_{f \in F} E[\rho_\tau(\ln(w) - X'\alpha(\tau) - \beta(\tau)S - f(X, Z))] \quad (3.4.7)$$

where F is the class of measurable functions of (X, Z) . In our empirical application, we restrict F to the values of Z_i , i.e., $f(X, Z) = Z'\hat{\gamma}$. To obtain an estimate for $\beta(\tau)$, we look

¹⁰ $u|X, Z \sim U(0, 1)$, i.e., for each S and S' given (X, S) : $U_S \sim U_{S'}$.

for a value $\hat{\beta}$ that makes the estimated coefficient on the instrumental variable $\hat{\gamma}(\beta, \tau)$ in equation (3.4.7) as close to zero as possible using a series of conventional quantile regression.

In practice, the IVQR estimator consists of a two-step procedure:¹¹ For a given value of $\beta^j(\tau)$, we first run the ordinary QR of $\ln(w) - \beta^j(\tau)S$ on X and Z to obtain the estimates $[\hat{\alpha}(\beta^j(\tau), \tau), \hat{\gamma}(\beta^j(\tau), \tau)]$. Second, we test $\hat{\gamma}(\beta^j(\tau), \tau) = 0$ and save the corresponding F -statistic, F_j . We then repeat these two steps for all the values in a pre-specified support for $\beta^j(\tau)$ and the value that minimizes the F -statistic is the IVQR estimator $\hat{\beta}(\tau)^{IVQR}$. Once we have $\hat{\beta}(\tau)^{IVQR}$, we retrieve the corresponding $\hat{\alpha}(\tau)$.¹²

The IVQR approach allows for an interpretation of the $\hat{\beta}(\tau)^{IVQR}$ as actual effects on individuals having fixed their level of unobserved heterogeneity at a given quantile. Therefore, the effect is not identified only for the set of individuals whose treatment is altered by switching the instrument from 0 to 1, as in the case of the IV quantile treatment estimator proposed by Abadie et al. (2002). Furthermore, the IVQR method puts no restriction of the form of the endogenous variables and instruments (i.e., they can be binary, discrete, or continuous).

3.4.3 Identification Strategy

Given the widely acknowledged endogeneity of educational attainment in the wage equation, finding valid instruments to control for this phenomenon is crucial. However, choosing suitable instruments remains a topic of great debate in the literature on returns to education (Arcand et al., 2005; Dickson, 2013; Heckman et al., 2006). In general, an ideal instrument should be correlated with educational attainment but uncorrelated with the unobserved determinants of the wage.

¹¹For further details, see Brunello et al. (2013) and Chernozhukov and Hansen (2008).

¹²To obtain the point estimates and standard errors, we use both the Stata command `ivqreg` and the Matlab function `invqr`, with almost no difference between the two approaches. The codes are publicly available at <http://faculty.chicagobooth.edu/christian.hansen/research/>

The literature on returns to education used several instruments for education: quarter of birth (Angrist and Krueger, 1991), early smoking habits (Evans and Montgomery, 1994), presence or gender of siblings (Butcher and Case, 1994), college proximity (Card, 1994), parental education (Harmon and Walker, 2000), and spouse education (Trostel et al., 2002). Over the past decade, the literature has been investigating educational reforms as a source of exogenous variation in educational attainment.¹³ In particular, changes in school-leaving age (Dickson, 2013; Harmon and Walker, 1999) and compulsory education expansions (Brunello et al., 2009, 2013; Fang et al., 2012) have been attracting research interest. Following this last strand of the literature, we exploit a major reform in the Swiss educational system to build our instruments and estimate the true (causal) effect of education on wages.

In Switzerland, the main responsibility for education and culture lies with the cantons, which loosely coordinate their work at the federal level. The 26 cantonal ministers of education together form a political body named the Swiss Conference of Cantonal Ministers of Education (EDK). The EDK is responsible for educational reforms, policies, and coordination at the national level. In 1970, the EDK proposed an important educational reform, with the aim of standardizing certain aspects of the Swiss education from compulsory school through high school. This reform became official on October 29, 1970. Previously, cantons had different compulsory school duration (seven, eight, or nine years) and different school year start (either spring or fall).

The reform set nine years of compulsory education for all cantons, and mandated that the school-year start in the fall. Given that some cantons were already in line with this reform, only about half had to change their education system. Moreover, cantons did not introduce the reforms immediately after 1970. They had time to adapt their education systems in the years following the agreement, with continuous feedback to the EDK on the reform status. Thus, we are able to keep track of the introduction of the

¹³For a recent study on the impact of educational reforms on educational attainment, see Braga et al. (2013).

reform in each canton. Additionally, to double-check the cantonal reform status, we also contacted each canton’s educational ministry. Appendix tables 3.7.2 and 3.7.3 give an overview of the reforms for each canton, the dates of their introduction in the canton (or not), and the ways in which the reforms modified (or not) the canton’s education system.

We use the compulsory education expansion as an instrument for years of education. The empirical literature suggests that postponing the allocation of pupils to tracks yields positive effects on average educational attainment, because students stay in school longer and drop out less (Braga et al., 2013). Similar to Brunello et al. (2009, 2013); Fang et al. (2012), we exploit the series of natural experiments created by the staggered implementation of Switzerland’s education reform as an instrument for estimating each individual’s completed years of schooling. This approach obviates the problem of endogeneity due to unobservable variables that are correlated with both education and wage. Compulsory schooling instrument might not work properly for individuals at the top of the distribution, because such high wage (ability) workers may be willing to acquire more schooling independent of the education expansion. The IVQR method does not allow to compute a first stage, but we can study the reduced-form effect to find out in which parts of the wage distribution we have identification from our instrument.

Given that the effective date of the shift in school-year start also constitutes a (small) exogenous change in years of education, we can also use this change as a second instrument for years of education completed. Pischke (2007) uses a similar approach for Germany, where a cohort experienced a shorter school year. In our case, however, the reform pertained all school levels from compulsory to high school, expanding the pool of “compliers.” Furthermore, the individuals affected by this second reform were different from the ones affected by the compulsory schooling expansion (different cantons and/or different year of introduction). Appendix tables 3.7.4 and 3.7.5 present TSLS and IVQR estimates for models using only the second instrument and over-identified models using both instruments.

3.5 Results and Discussion

3.5.1 Causal Returns to Education over Wage Distribution

Table 3.5.1 shows the regression outputs for model (3.4.1), which focuses on the returns to education. Mean regression (column 1, table 3.5.1) estimates a return to education of 6.7 percent, which indicates that wages rise by almost seven percent on average with each extra year of education. The effect is highly significant and not far from the few previous studies on returns to education in Switzerland, which estimate returns of about 7-8 percent (Weber and Wolter, 1999).

When we allow for heterogeneous effects of education on wage, an interesting picture emerges. QR estimates (columns 2-6, table 3.5.1) show that returns to education increase over the quantiles of the wage distribution. The return to education is 3.9 percent at the bottom decile, increasing to 6.9 percent at the median ($\tau = 0.5$), and reaching 8.9 percent at the top decile of the wage distribution. These results already underline that average effects hide useful information about the rest of the distribution: Further emphasizing the heterogeneous effects of education on wage, Figure 3.5.1 reports the quantile-specific returns to education from $\tau = 0.1$ to $\tau = 0.9$. Again, we observe increasing returns over the wage distribution. Our estimated return patterns over the wage distribution is very similar to those found by the literature for other countries (Fasih et al., 2012; Harmon et al., 2003; Hartog et al., 2001).

Until now, however, we did not consider the endogeneity of years of education in our empirical analysis. Therefore, Table 3.5.2 presents TSLS estimates of model (3.4.1). As an instrument for years of education we use the expansion in compulsory education that took place in some cantons after 1970. The returns to education estimated by TSLS are slightly higher than OLS estimates, with a point estimate (standard error) of 9.9 percent (0.019). This result is typical in the literature on returns to education and is usually motivated by measurement error in the education variables (Card, 2001) and

Table 3.5.1: RETURNS TO EDUCATION, OLS AND QR ESTIMATES

Variables	OLS (1)	1 st Decile (2)	3 rd Decile (3)	Median (4)	7 th Decile (5)	9 th Decile (6)
Years of education	0.067** (0.001)	0.039** (0.001)	0.060** (0.001)	0.069** (0.001)	0.076** (0.001)	0.089** (0.001)
Age	0.053** (0.001)	0.037** (0.002)	0.042** (0.001)	0.048** (0.001)	0.054** (0.001)	0.067** (0.002)
Age ² /100	-0.051** (0.002)	-0.039** (0.003)	-0.042** (0.001)	-0.046** (0.002)	-0.051** (0.002)	-0.064** (0.003)
Constant	9.118** (0.029)	9.474** (0.050)	9.287** (0.022)	9.163** (0.024)	9.056** (0.027)	8.801** (0.050)
Year fixed effects	YES	YES	YES	YES	YES	YES
(Pseudo) R ²	0.292	0.060	0.149	0.206	0.239	0.232
N	34,744	34,744	34,744	34,744	34,744	34,744

Notes: ** $p < 0.01$, * $p < 0.05$, † $p < 0.10$. Robust standard errors are in parentheses. The dependent variable is the natural logarithm of annual wage.

Swiss Labor Force Survey, Authors' calculations.

Figure 3.5.1: Returns to Education, Quantile Regression Estimates

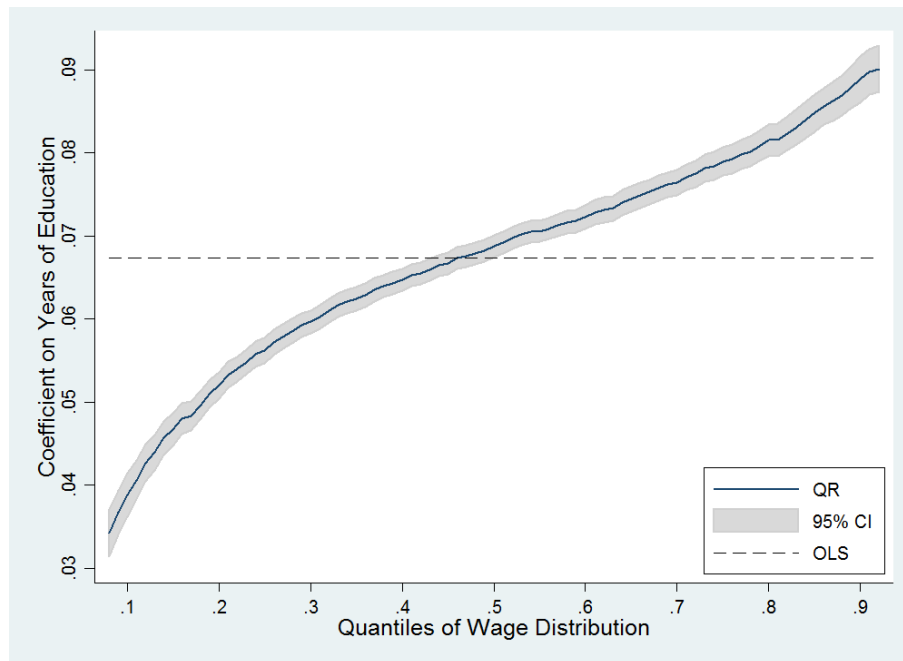


Table 3.5.2: RETURNS TO EDUCATION, TSLS ESTIMATES

Variables	OLS (1)	Reduced Form (2)	First Stage (3)	Second Stage (4)
Years of education	0.067** (0.001)			0.099** (0.019)
Age	0.053** (0.001)	0.065** (0.002)	0.174** (0.011)	0.048** (0.003)
Age ² /100	-0.051** (0.002)	-0.066** (0.002)	-0.211** (0.014)	-0.046** (0.004)
Constant	9.118** (0.029)	9.748** (0.033)	10.192** (0.228)	8.725** (0.214)
IV-Education expansion		0.034** (0.007)	0.346** (0.054)	
Year fixed effects	YES	YES	YES	YES
R ²	0.292	0.095	0.018	0.249
N	34,744	34,744	34,744	34,744
Test for excluded instruments				
<i>F</i> -statistic			40.75**	
Under-identification test				
Kleibergen-Paap LM-statistic				39.90**

Notes: ** $p < 0.01$, * $p < 0.05$, † $p < 0.10$. Robust standard errors are in parentheses. In columns (1), (2), and (4) the dependent variable is the natural logarithm of annual wage, in column (3) the dependent variable is years of education.

Swiss Labor Force Survey, Authors' calculations.

local average treatment effects (Imbens and Angrist, 1994; Imbens, 2010). The coefficient on the instrumental variable in both the reduced form and first stage has the expected sign and is highly significant. In the first stage model (column 3 of table 3.5.2), our instrument has a positive and significant effect on years of education. This finding is in line with the expectations discussed in section 3.4.3, and is consistent with studies that use similar instruments (Braga et al., 2013; Brunello et al., 2009, 2013; Fang et al., 2012). In our specific case, the reforms increased educational attainment by one third of a year on average, whereas previous studies estimated an effect of about half a year. The test for excluded instruments has an F -statistic of 40.75, which is well beyond the accepted standard of 10 (Staiger and Stock, 1997). We are therefore confident about the strength of the instrumental variable. We also reject the null hypotheses of under-identification for our instrument (Kleibergen-Paap statistic).

Table 3.5.3 shows the IVQR estimates of model (3.4.1). With this regression analysis we causally estimate the impact of education on wage at a given quantile of the wage distribution. As in the standard QR analysis, IVQR results also show that the returns to schooling are heterogeneous over the wage distribution. However, the shapes of the estimated returns over the quantiles are very different: the causal effect of education is larger in the lower parts of the wage distribution, and the effect becomes small or insignificant at the top.

Specifically, the return to education estimated by IVQR is 18.3 percent at the first decile, decreasing to 9.6 percent at the median, and going down to an insignificant 1.6 percent at the last decile of the wage distribution. These results indicate that the largest gains to additional years of education accrue to individuals at the low end of the wage distribution. Figure 3.5.2 provides a graphical illustration of these results from $\tau = 0.1$ to $\tau = 0.9$, with a quantile interval of 0.05. As the reduced-form quantile IV approach produces qualitatively similar point estimates and distributional patterns to the structural IVQR approach, we are confident that our substantive results are not sensitive to the estimation procedure (Autor et al., 2012) and we see how important it is to study the quantiles and not only averages.

A look at the reduced-form effects might explain the drop in returns at the top of the wage distribution. As Figure 3.5.2 indicates, for top earners we do not have a reduced-form effect, making it impossible to compute the respective instrumental variable estimate. Therefore, the drop in return in the top decile is due to a loss of identification rather than a zero causal effect of one additional year of education. This finding is consistent with our discussion of subsection 3.4.3, in which we argued that our instrument would not work properly for individuals at the top of the distribution.

The IVQR estimates are also consistent with the theoretical expectations we formulated previously. As the quantile index τ can be viewed as a measure of unobserved individual ability, the IVQR results are in line with the argument that individuals ac-

Table 3.5.3: RETURNS TO EDUCATION, TSLS AND IVQR ESTIMATES

Variables	TSLS	1 st Decile	3 rd Decile	Median	7 th Decile	9 th Decile
	(1)	(2)	(3)	(4)	(5)	(6)
Years of education	0.099** (0.019)	0.183** (0.017)	0.169** (0.013)	0.096** (0.010)	0.066** (0.010)	0.016 (0.044)
Age	0.048** (0.003)	0.005 (0.004)	0.026** (0.003)	0.046** (0.002)	0.055** (0.002)	0.100** (0.007)
Age ² /100	-0.046** (0.004)	0.005 (0.004)	-0.022** (0.003)	-0.045** (0.002)	-0.052** (0.002)	-0.097** (0.009)
Constant	8.725** (0.199)	7.781** (0.076)	8.048** (0.055)	8.828** (0.042)	9.178** (0.043)	9.166** (0.476)
Year fixed effects	YES	YES	YES	YES	YES	YES
Reduced form effect	0.034** (0.007)	0.063** (0.012)	0.046** (0.006)	0.037** (0.006)	0.035** (0.007)	-0.001 (0.011)
<i>N</i>	34,744	34,744	34,744	34,744	34,744	34,744

Notes: ** $p < 0.01$, * $p < 0.05$, † $p < 0.10$. Robust standard errors are in parentheses. The dependent variable is the natural logarithm of annual wage.

Swiss Labor Force Survey, Authors' calculations.

quire education up to the point where the cost equals the rate of return and where costs depend negatively on ability (Card, 1999). In this setting, we would expect the returns to education to be decreasing in ability, with the lower-ability individuals having the highest return to education—which is exactly the pattern estimated by IVQR. Moreover, interpreting the quantile index as an ability measure is also consistent with the notion that individuals with higher ability are likely to generate higher wages regardless of their educational level. Conversely, individuals with lower unobserved ability would gain more from the training provided by formal education. Our estimates suggest that higher-ability individuals indeed have higher wages, but the slope of their wage-education profile is flatter than that for lower-ability individuals.

As robustness check, we also use a different instrument (shift in school-year start) and a combination of two instruments (compulsory education expansion and shift in school-year start), which does not largely affect the estimated returns to education (see appendix tables 3.7.4 and 3.7.5). However, with multiple instruments we have a gain in the precision of the education coefficient and we can test for over-identification. The

Figure 3.5.2: Returns to Education, IVQR Estimates



p -value of the Hansen statistic is always not significant, indicating that we cannot reject the null hypothesis that all our model assumptions are fulfilled—including the validity of the instruments. This constitutes an important validity check for our estimates, which we pass at all instances.

3.5.2 Heterogeneous Returns Between and Within Types of Education

We now focus on the comparison between educational types. Table 3.5.4 gives an overview of the OLS and QR estimates of model (3.4.2). Column 1 of Table 3.5.4 presents OLS regressions, which estimate a return to vocational education of 6.8 percent and a return to academic education of 7.1 percent. These coefficients gather the effect of an extra year of vocational (academic) education on wage, filtering out the effect of compulsory schooling. By performing an F -test, we reject the null hypothesis of equal coefficients ($p = 0.00$), i.e., at the mean, the effect of one additional year of academic education on wage is larger than the effect on one additional year of vocational education. The

question, however, is whether modeling on average loses some important features of this comparison. Therefore, we bring the discussion into a distributional framework.

Columns 2-6 of Table 3.5.4 present the QR estimates for model (3.4.2) at various quantiles of the wage distribution. The first result is that, as in Table 3.5.1, returns to both vocational and academic education are increasing in the quantiles of the wage distribution. However, the increasing pattern and the magnitude of the estimated effects are significantly different. At the lower quantiles of the wage distribution, vocational education has a statistically significant return premium in comparison to academic education. From the fourth decile on, the situation is reversed: Academic education has higher returns for one additional year of schooling. Thus, in the upper part of the wage distribution, academic education brings a significant premium compared to vocational education.

In particular, at the bottom decile, the return to one extra year of vocational education is 5.0 percent, whereas the return to one additional year of academic education is only 4.1 percent. We reject the null hypothesis of equal coefficients at each level of significance ($p = 0.00$). At the third decile the situation is different, with an estimated return of about 6.4 percent for both academic and vocational tracks ($p = 0.66$). At the median, the returns to vocational and academic educations are 6.9 percent and 7.3 percent, respectively. Similarly to OLS, at the median we reject the null hypothesis of equal coefficients, with a p -value of 0.00. At the top decile, academic education brings a return of 9.6 percent, while vocational education has an estimated return of 8.3 percent. The difference between the estimated coefficients is statistically significant ($p = 0.00$). Figure 3.5.3 provides graphical support complementing the Table 3.5.4 results that we just discussed, comparing OLS estimates with QR estimates across the entire wage distribution, estimated for all quantiles from $\tau = 0.1$ to $\tau = 0.9$.

For a better understanding of the academic premium, we rewrite equation (3.4.2) as a function of the difference between the two educational tracks, with vocational ed-

Table 3.5.4: RETURNS TO VOCATIONAL AND ACADEMIC EDUCATION, OLS AND QR

Variables	OLS (1)	1 st Decile (2)	3 rd Decile (3)	Median (4)	7 th Decile (5)	9 th Decile (6)
Vocational education	0.068** (0.001)	0.050** (0.002)	0.064** (0.001)	0.069** (0.001)	0.075** (0.001)	0.083** (0.002)
Academic education	0.071** (0.001)	0.041** (0.002)	0.064** (0.001)	0.073** (0.001)	0.082** (0.001)	0.096** (0.001)
Compulsory education	-0.032** (0.003)	-0.062** (0.006)	-0.038** (0.003)	-0.028** (0.003)	-0.018** (0.004)	-0.015* (0.007)
Age	0.045** (0.001)	0.029** (0.003)	0.035** (0.001)	0.041** (0.001)	0.046** (0.001)	0.058** (0.002)
Age ² /100	-0.043** (0.002)	-0.031** (0.003)	-0.035** (0.002)	-0.039** (0.002)	-0.043** (0.002)	-0.053** (0.003)
Constant	10.133** (0.045)	10.474** (0.081)	10.259** (0.042)	10.145** (0.039)	10.043** (0.046)	9.921** (0.081)
Year fixed effects	YES	YES	YES	YES	YES	YES
(Pseudo) R ²	0.309	0.072	0.163	0.220	0.250	0.243
F-statistic $\hat{\beta}_V = \hat{\beta}_A$	11.71**	23.63**	0.190	22.94**	51.06**	62.09**
N	34,744	34,744	34,744	34,744	34,744	34,744

Notes: ** $p < 0.01$, * $p < 0.05$, $\dagger p < 0.10$. Robust standard errors are in parentheses. The dependent variable is the natural logarithm of annual wage.

Swiss Labor Force Survey, Authors' calculations.

Figure 3.5.3: Returns to Vocational and Academic Education, Quantile Regression Estimates

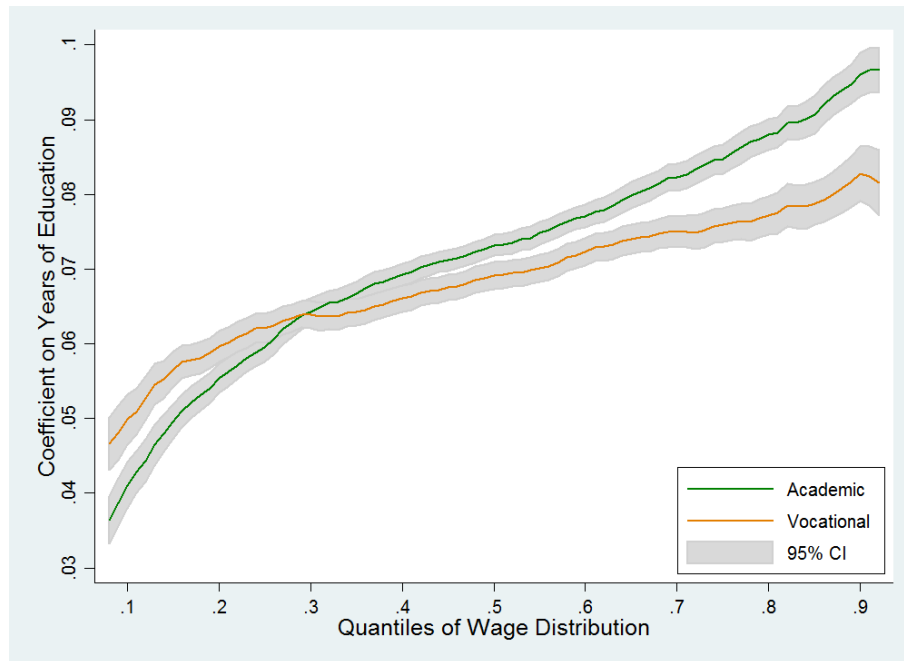
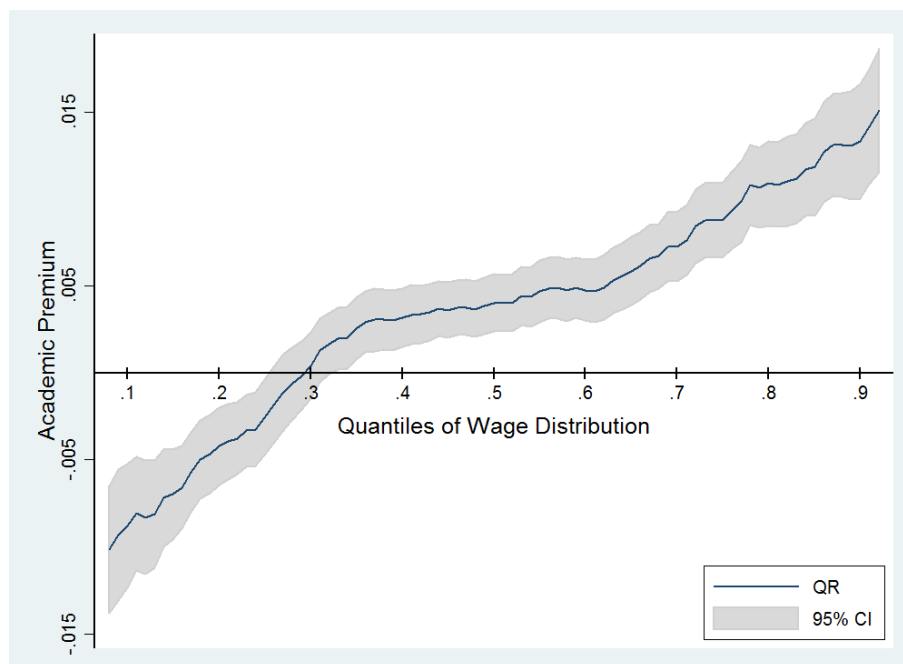


Figure 3.5.4: Academic Education Premium, Quantile Regression Estimates



ucation as the reference category. While doing so prevents us from seeing the pattern of vocational and academic educations separately, the transformation allows estimating confidence intervals for the academic premium in comparison to vocational. Figure 3.5.4 plots the academic premium over the wage distribution, along with its 95 percent confidence intervals.

One potential explanation for these results is the skill formation of vocational and academic education. While the vocational education system provides a set of skills that are specific to the job that the apprentices are learning (Busemeyer and Trampusch, 2012), in academic education the exploitation of the acquired skills strongly depends on whether or not the workers are using them in the labor market (Dearden et al., 2002). In addition, vocational education is likely a better fit for students at the lower part of the wage distribution, because those students learn contents that better match and complement their innate abilities (Rosenbaum and Rosenbaum, 2013). As a consequence of this skill formation and sorting mechanism, at the lower quantiles of the wage distribution, vocational education brings a return premium because individuals with an academic education in this part of the distribution have a relative disadvantage in the job they are

performing. Conversely, at some point in the wage distribution (in our data $\tau = 0.4$) academic education, as opposed to vocational education, starts generating a return premium because workers have the capacity of fully exploiting their skills in the labor market.

Given that we are more interested in the presence of heterogeneity and because we did not find appropriate instrumental variables for both academic and vocational education, we do not claim that the estimated effects in the between-within path comparison are causal. We nevertheless performed some simple two-stage quantile regressions following the approach of Chen and Portnoy (1996), based on early work by Powell (1983). Regression outputs are depicted in appendix Table 3.7.6. We use a dummy that equals one if the canton of residence has a university as an instrument for academic education,¹⁴ whereas for vocational education we exploit regional variation in preference for vocational education compared to academic education.¹⁵ Although we do want to put too much emphasis on these estimations because the instruments we use are only *arguably* exogenous and because the local average treatment effects are difficult to interpret, it is important to notice that we find—again—lower returns to academic education at the bottom of the wage distribution and a return premium for academic education in the upper part of the wage distribution. Thus it appears that, for the comparison of academic and vocational education, the qualitative results do not change systematically if we take endogeneity into account.

3.6 Conclusions

This chapter presents evidence of heterogeneous returns to education over the wage distribution. We use instrumental variable quantile regression and data from the Swiss Labor Force Survey to isolate the causal link between education and wages at different quantiles of the conditional distribution of wages. Our results provide significant evidence that no

¹⁴Dee (2004) and Card (1993) use a similar approach.

¹⁵As in Rupietta and Backes-Gellner (2012).

unique causal effect of schooling exists and that for individuals the effect may be above or below the average returns extensively documented using OLS or TSLS, depending on their position in the wage distribution and their unobservable wage determinants. As standard economic theory suggests that workers are paid according to their productivity, we can interpret the individual's position in the wage distribution as a proxy for ability. This assumption—a common one in the literature (Card, 2012)—provides our results with additional implications.

In particular, while ordinary QR results indicate that returns to education are increasing in the quantile index, once we take the endogeneity of schooling into account, we instead observe higher returns at lower quantiles of the wage distribution. Interpreting the quantile index as a measure of unobserved ability, our findings suggest that less able individuals profit more from one additional year of education. While higher-ability individuals have on average higher wages, the slope of their wage-education profile is flatter than that for lower-ability individuals. This finding indicates, as discussed by Ashenfelter and Rouse (1998), that more able individuals acquire more schooling because they face lower marginal costs, not because they receive higher marginal benefits.

From a methodological point of view, one noteworthy result of our analysis is that a reduced-form quantile IV approach, akin to TSLS, produces qualitatively similar estimates to the structural IVQR approach, which is based on stronger assumptions. The comparability of these estimates indicates that our core results are not sensitive to the estimation procedure.

We also investigate the potential heterogeneity in the returns within and between different educational paths. Exploiting the uniqueness of the Swiss educational system, we complement the existing literature by confirming that, at the mean, academic education brings higher returns. However, if we examine the returns over the wage distribution, we observe two relevant—and until now unknown—features of the returns to vocational and academic education. First, we reveal significant heterogeneity within each educa-

tional path, with both vocational and academic education presenting increasing returns over the wage distribution. Second, a comparison between the two tracks shows that academic education does not always yield higher returns. In the upper part of the wage distribution, individuals with an academic background have higher returns than individuals with a vocational background. However, at lower quantiles of the wage distribution, vocational education brings higher returns than academic education. These results imply that answering the question of whether academic education yields higher labor market returns than vocational education is not as easy as it might have once appeared from descriptive statistics or mean regression. Indeed, the answer depends on the individual's position in the conditional wage distribution.

Our work can be extended in a number of ways. First, analyzing the evolution over time of the quantile returns to education, and what impact the returns have on the structure of wages, would be valuable. According to our results, education should have an inequality-reducing effect over time, because individuals with lower ability (i.e., those at the lower quantiles of the wage distribution) appear to profit more from formal education. However, such inquiry is complicated by the likelihood that the endogeneity and measurement error biases change over time.

Second, in line with several cross-country studies conducted for example by Martins and Pereira (2004) and by Trostel et al. (2002), researchers and policymakers might use an international comparison to study how the causal returns to education change with different wage distributions and education systems. Third, researchers could explore the potential non-linear relationship between education and wages by allowing the returns to differ not only between educational paths but also across education levels, as, for example, in Buchinsky (1994); Hartog et al. (2001).

A fourth, and compelling, extension to our work would be evaluating the impact of changes in the distribution of education on quantiles of the unconditional (marginal) distribution of wages. Doing so would help estimate the effect of one additional year of

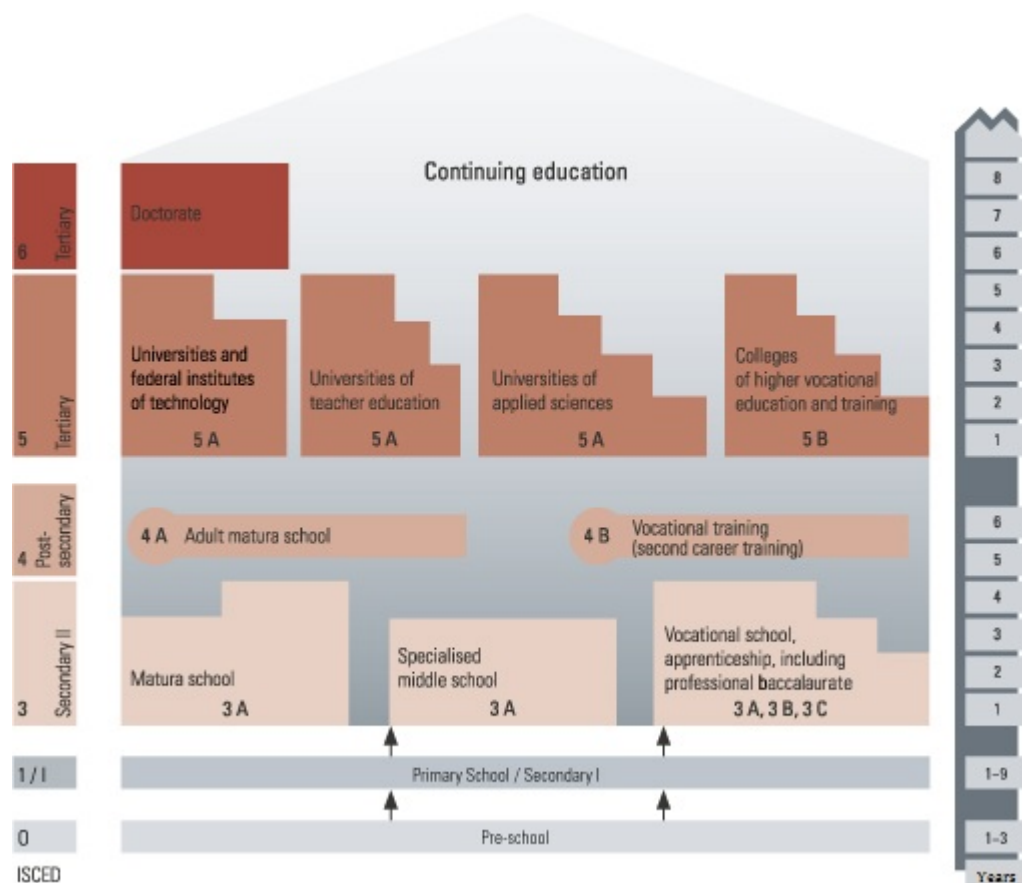
schooling on the entire wage distribution, not only at a given quantile. However, to shed light on this topic, we would need an adaptation of the unconditional QR approach (Firpo et al., 2009) to instrumental variables estimation—an adaptation not yet available, as we discussed in the previous chapter.

This chapter shows that typical estimates of the mean return to education provide a relatively incomplete characterization of the impact of education on labor market outcomes and thus constitute a weak guide for public policy. Similarly, distributional analyses using ordinary QR also constitute an inappropriate tool for describing the true impact of education on wages, because they do not control for unobserved heterogeneity. Our results suggest that the true impact of education on the distribution of wages is highly heterogeneous, and we empirically support the argument that formal education partially compensates for differences in innate abilities and early life conditions.

3.7 Appendix

3.7.1 The Swiss Education System

Figure 3.7.1: The Swiss Education System (Source: Swiss Federal Statistical Office)



3.7.2 Analytic Sample

Table 3.7.1: SAMPLE CONSTRUCTION

Initial sample (SLFS 2000–2009)	160,925
Males	74,871
Fully employed	47,347
Age between 18 and 60	44,670
Not in education or gap year	42,612
Wage not missing	35,095
99 percent of wage distribution	34,744
Analytic sample	34,744

Notes: Swiss Labor Force Survey, Authors' calculations. Column 1 presents the variables considered to create the sample, while column 2 shows the number of observations left after each sample restriction.

3.7.3 Summary of the Reform of 1970

Table 3.7.2: COMPULSORY EDUCATION EXPANSION

Canton	Entry Age (1)	Reform (2)	Year (3)	Before (4)	After (5)	First Cohort (6)
Zürich	6	Yes	1977	8	9	1971
Bern	6	No		9	9	
Luzern	6	Yes	1985	8	9	1979
Uri	7	Yes	1977	7	9	1970
Schwyz	7	Yes	1992	7	9	1985
Obwalden	7	Yes	1992	7	9	1985
Nidwalden	6	Yes	1992	7	9	1986
Glarus	6	Yes	1983	8	9	1977
Zug	7	Yes	1990	8	9	1983
Fribourg	7	No		9	9	
Solothurn	7	Yes	1970	8	9	1963
Basel-Stadt	6	No		9	9	
Basel-Land	6	Yes	1980	8	9	1974
Schaffhausen	6	Yes	1982	8	9	1976
Appenzell A.	6	Yes	1981	8	9	1975
Appenzell I.	6	Yes	1984	7	9	1978
St. Gallen	6	Yes	1983	8	9	1977
Graubünden	7	No		9	9	
Aargau	7	Yes	1982	8	9	1975
Thurgau	6	Yes	1980	8	9	1974
Ticino	6	No		9	9	
Vaud	7	No		9	9	
Valais	7	Yes	1987	8	9	1980
Neuchâtel	6	No		9	9	
Genève	6	No		9	9	
Jura	6	No		9	9	

Notes: Authors' research and calculations. Column 1 shows the entry age in compulsory schooling, column 2 indicates whether a canton had to reform its educational system, and, if so, in which year (column 3). Columns 4 and 5 show the years of compulsory schooling before and after the reform, and column 6 indicates the first cohort affected by the education reform.

Table 3.7.3: CHANGES IN SCHOOL-YEAR START

Canton	Entry Age (1)	Reform (2)	Year (3)	Before (4)	After (5)	First Cohort (6)
Zürich	6	Yes	1989	Spring	Fall	1974
Bern	6	Yes	1989	Spring	Fall	1974
Luzern	6	No		Fall	Fall	
Uri	7	No		Fall	Fall	
Schwyz	7	Yes	1989	Spring	Fall	1975
Obwalden	7	No		Fall	Fall	
Nidwalden	6	No		Fall	Fall	
Glarus	6	Yes	1989	Spring	Fall	1975
Zug	7	Yes	1973	Spring	Fall	1958
Fribourg	7	No		Fall	Fall	
Solothurn	7	Yes	1989	Spring	Fall	1973
Basel-Stadt	6	Yes	1989	Spring	Fall	1974
Basel-Land	6	Yes	1989	Spring	Fall	1975
Schaffhausen	6	Yes	1989	Spring	Fall	1975
Appenzell A.	6	Yes	1989	Spring	Fall	1975
Appenzell I.	6	Yes	1989	Spring	Fall	1976
St. Gallen	6	Yes	1989	Spring	Fall	1975
Graubünden	7	No		Fall	Fall	
Aargau	7	Yes	1989	Spring	Fall	1974
Thurgau	6	Yes	1989	Spring	Fall	1974
Ticino	6	No		Fall	Fall	
Vaud	7	Yes	1973	Spring	Fall	1957
Valais	7	No		Fall	Fall	
Neuchâtel	6	Yes	1973	Spring	Fall	1958
Genève	6	No		Fall	Fall	
Jura	6	Yes	1989	Spring	Fall	1974

Notes: Authors' research and calculations. Column 1 shows the entry age in compulsory schooling, column 2 indicates whether a canton had to reform its educational system, and, if so, in which year (column 3). Columns 4 and 5 show the season of school start before and after the reform, and column 6 indicates the first cohort affected by the education reform.

3.7.4 Alternative Instrument and Over-identified Models

Table 3.7.4: RETURNS TO EDUCATION, TSLS ESTIMATES

Variables	First Stage (1)	Second Stage (2)	First Stage (3)	Second Stage (4)	First Stage (5)	Second Stage (6)
Years of education		0.104** (0.026)		0.101** (0.017)		0.102** (0.014)
Age	0.166** (0.011)	0.047** (0.004)	0.185** (0.011)	0.048** (0.003)	0.179** (0.011)	0.048** (0.003)
Age ² /100	-0.201** (0.013)	-0.045** (0.005)	-0.221** (0.014)	-0.045** (0.004)	-0.212** (0.014)	-0.045** (0.003)
Constant	10.348** (0.223)	8.669** (0.288)	9.876** (0.240)	8.703** (0.188)	9.978** (0.241)	8.696** (0.159)
IV_1 -Education expansion			0.302** (0.055)	0.620** (0.083)		
IV_2 -Shift in school start	0.221** (0.043)		0.178** (0.044)	0.267** (0.047)		
$IV_1 \cdot IV_2$				-0.539** (0.104)		
Year fixed effects	YES	YES	YES	YES	YES	YES
R ²	0.018	0.233	0.019	0.243	0.019	0.241
N	34,744	34,744	34,744	34,744	34,744	34,744
Test for excluded instruments						
F -statistic	26.69**		28.37**		26.53**	
Under-identification test						
Kleibergen-Paap LM-statistic		26.58**		55.91**		77.95**
Over-identification Test						
Hansen J-statistic (p -value)				0.860		0.982

Notes: ** $p < 0.01$, * $p < 0.05$, † $p < 0.10$. Robust standard errors are in parentheses. In odd columns the dependent variable is years of education, in even columns the dependent variable is the natural logarithm of annual wage.
Swiss Labor Force Survey, Authors' calculations.

Table 3.7.5: OVER-IDENTIFIED IVQR

Variables	25 th Percentile (1)	Median (2)	75 th Percentile (3)	25 th Percentile (4)	Median (5)	75 th Percentile (6)
Years of education	0.178** (0.032)	0.090** (0.023)	0.082** (0.025)	0.164** (0.025)	0.085** (0.019)	0.092** (0.021)
Age	0.020** (0.005)	0.047** (0.004)	0.056** (0.004)	0.023** (0.005)	0.047** (0.003)	0.055** (0.004)
Age ² /100	-0.014* (0.007)	-0.045** (0.005)	-0.053** (0.005)	-0.017** (0.006)	-0.045** (0.004)	-0.051** (0.005)
Constant	8.143** (0.338)	8.910** (0.250)	8.979** (0.270)	8.124** (0.281)	8.972** (0.207)	8.882** (0.228)
Year fixed effects	YES	YES	YES	YES	YES	YES
IV ₁ -Education expansion	X	X	X	X	X	X
IV ₂ -Shift in school start	X	X	X	X	X	X
IV ₁ · IV ₂				X	X	X
N	34,744	34,744	34,744	34,744	34,744	34,744

Notes: ** $p < 0.01$, * $p < 0.05$, † $p < 0.10$. Robust standard errors are in parentheses. The dependent variable is the natural logarithm of annual wage.

Swiss Labor Force Survey, Authors' calculations.

3.7.5 Returns to Vocational and Academic Education Instrumented

Table 3.7.6: RETURNS TO VOCATIONAL AND ACADEMIC EDUCATION, TSLS AND TWO-STAGE QR

Variables	TSLS (1)	25 th Percentile (2)	Median (3)	75 th Percentile (4)
Vocational education	0.128** (0.006)	0.139** (0.004)	0.161** (0.006)	0.127** (0.008)
Academic education	0.130** (0.012)	0.091** (0.011)	0.138** (0.013)	0.162** (0.016)
Compulsory education	-0.032** (0.005)	-0.033** (0.004)	-0.027** (0.005)	-0.026** (0.007)
Age	0.032** (0.003)	0.021** (0.003)	0.025** (0.003)	0.043** (0.004)
Age ² /100	-0.028** (0.004)	-0.020** (0.003)	-0.022** (0.004)	-0.036** (0.005)
Constant	10.051** (0.070)	10.148** (0.061)	10.026** (0.074)	9.871** (0.093)
Year fixed effects	YES	YES	YES	YES
IV_{ACA} –University canton	X	X	X	X
IV_{VET} –German canton	X	X	X	X
Test for excluded IVs				
F -statistic	56.29**			
Under-identification test				
Kleibergen-Paap statistic	111.14**			
(Pseudo) R^2	0.163	0.060	0.072	0.087
F -statistic $\hat{\beta}_V = \hat{\beta}_A$	0.08	32.62**	5.48*	8.32**
N	34,744	34,744	34,744	34,744

Notes: ** $p < 0.01$, * $p < 0.05$, † $p < 0.10$. Robust standard errors are in parentheses. The dependent variable is the natural logarithm of annual wage.

Swiss Labor Force Survey, Authors' calculations.

Chapter 4

The Effects of Involuntary Separations on Wage Trajectories

[A version of this paper has been submitted to the *German Economic Review*]

4.1 Introduction

In the last two decades, concerns about the plight of job loss have been a relevant issue for both researchers and policymakers (Couch and Placzek, 2010; Hijzen et al., 2010; White, 2010). Previous studies analyzing involuntary separations have shown that this type of job loss implies monetary costs in terms of immediate wage reductions (Curti, 1998; Monks and Pizer, 1998) and non-monetary costs in terms of re-employment conditions (Farber, 2010; Polsky, 1999). It remains unclear, however, whether these wage reductions are long lasting.

To fill this gap, the purpose of this paper is to shed light on the effects of involuntary

separations on wages in the long term. To do this, we first estimate the earnings losses of those who find re-employment over several years after an involuntary job loss. Second, we are not only interested in measuring earnings losses of those that are re-employed, but we are also interested in the total foregone earnings or the total productivity loss of those that experience unemployment spells. Therefore, in a second step we also include unemployed workers with zero earnings in our analysis. To do so, we have to use a new empirical approach: a Poisson pseudo-maximum-likelihood estimator similar to the one applied to gravity models (Santos Silva and Tenreyro, 2006). Including these “zeros” is important for at least two reasons. First, by considering the zeros, we are no longer only measuring firm-specific loss of human capital (Addison and Portugal, 1989; Carrington, 1993; Neal, 1995) but instead estimating the total foregone productivity caused by an involuntary separation.¹ The unanswered question here is whether these productivity losses are also long-lasting or whether they affect only the year of separation. Second, from the methodological side, it reduces the selection bias due to the exclusion of one portion of the population, as underlined by von Wachter et al. (2008).

We further deepen our analysis and investigate the determinants of an involuntary separation, asking whether quantity and type of education can act as a protection against job loss. Previous literature (Kettunen, 1997; Mincer, 1991) suggests that unemployment incidences are lower among highly educated workers. For example, Polsky (1999) shows that having either a college degree or at least some college is negatively correlated with job separation. In this chapter, we not only focus on educational level (i.e., years of schooling) but we also distinguish the type of education that people received in those years (i.e., academic or vocational).

Our data comes from the Swiss Labor Force Survey from 1996 to 2009, a rotating panel representative of the adult population permanently living in Switzerland. The Swiss Labor Force Survey is produced annually by the Swiss Federal Statistical Office,

¹For example, if we did not include these zeros, we would analyze the wage losses of those workers who directly find a job after separation, excluding all individuals experiencing an unemployment spell.

and the main advantage of this data source is that we can fully observe the reason of each job separation. Therefore, we can distinguish between workers who separated involuntarily from those who separated voluntarily or left for other reasons (e.g., injury, working conditions, or personal issues) and obtain valid estimates of the wage trajectories following an individual before, during, and after an involuntary job loss. Furthermore, we can compare the earnings patterns for different types of separation to test whether the consequences of self-reported reasons for separation are in line with our theoretical expectations.

Our results show that the wage losses following an involuntary separation are significant and long-lasting. Separated workers suffer from an immediate loss of about 10 percent, a loss remaining statistically significant at least for four years after separation at about 11 percent. If we include individuals with zero earnings to get an estimate for total productivity loss, we find losses of 40 percent in the year of separation and a long-term loss of about 19 percent four years after separation. These larger estimated losses can be seen as an indicator of the total productivity loss caused by an involuntary separation. This is because they comprise of zero labor market productivity during unemployment and the loss of firm-specific human capital that is reflected by the lower wages after re-employment (Addison and Portugal, 1989).

Compared to other reasons for separation, the earnings loss pattern is unique for involuntary separations, because no other type of separation implies such permanent scarring. Regarding the role of education, we find that tertiary education—either academic or vocational—plays a major role in reducing the risk of job loss.

The remainder of this chapter proceeds as follows: section 4.2 gives an overview of the theoretical framework and empirical literature related to our research questions; section 4.3 introduces the data set and gives some descriptive statistics; section 4.4 introduces the equations to be estimated and shows the econometric approaches in detail; section 4.5 presents the results, and section 4.6 concludes.

4.2 Theory and Empirical Background

Monks and Pizer (1998) present early evidence of involuntary separations. They use data from the U.S. National Longitudinal Surveys to estimate the increase in the probability of changing jobs from 1971 to 1990, and identify how much of this change was voluntary. Overall, they find a positive trend in the probability of job turnover of about 13 percent, which can be decomposed into an insignificant increase in the probability of quitting voluntarily and a significant 6.8-percent increase in the probability of involuntary separation.

Both confirming Monks and Pizer's results and extending the literature, Polsky (1999) examines the consequences of job loss between 1976-81 and 1986-91. In general, he finds stability in the incidence of job separation but a statistically significant increase in the incidence of involuntary job loss relative to quitting. Polsky also shows that the consequences of involuntary separation worsened over time: The re-employment rate of workers who experienced involuntary job loss dropped from 67 percent to 62 percent in 15 years. Moreover, among those who found new jobs, the odds of receiving a considerable wage cut rose from 9 percent to 17 percent during the same period. When these earnings losses start and whether they are long lasting remains uninvestigated.

The literature on earnings losses focuses primarily on displaced workers, who have been involuntarily separated due to plant closings or mass lay-off.² This focus on displacements due to plant closures and mass layoffs is meant to avoid selection problems. Displacements are considered exogenous and usually affect the entire workforce to the same extent. However, the selection problem then remains at the firm level, because establishments that have to close are likely to be in the low-performing tail of the distribution of plants and most likely also workers (Hijzen et al., 2010). Therefore, we decide to use individually reported reasons for separation.

²The U.S. Bureau of Labor Statistics definition of a displaced worker is [s]omeone at least 20 years old, with at least three years of tenure on a job, who lost that job due to slack work, abolition of a position or shift, or plant closing or relocation.

Regarding the persistence of earnings losses after separation, Ruhm (1991) introduces the concept of “long-lasting scars” following job displacements. Using household data from the Michigan Panel Study of Income Dynamics, he shows that four years after displacement, displaced workers continue to earn 10-13 percent less than their non-displaced counterparts. The methodological standard for the earnings losses literature is Jacobson et al. (1993), who use administrative data³ combining workers’ earnings histories with information about their firms to estimate the magnitude and temporal pattern of displaced workers’ earnings losses. They find that high-tenure workers separating from distressed firms suffer an immediate loss of more than 40 percent and a long-term loss of about 25 percent per year.

Couch and Placzek (2010) conduct a revisited analysis of displaced workers’ earnings losses. They argue that past estimates and the size of those reductions vary strongly with the type of data used and the business cycle conditions, demonstrating that under ordinary economic conditions, earnings losses are smaller than the estimations of Jacobson et al. White (2010) obtains similar results, finding the typical displaced worker realizing total long-term losses of USD 34,065, equivalent to an 11-percent loss compared to the earnings of similar non-displaced workers.

In contrast to the U.S. evidence, most European studies find displacement losses rather small and not always long-lasting (Couch, 2001; Eliason and Storrie, 2006; Hijzen et al., 2010). As European countries usually have stronger labor market regulations in which wages are bargained between employers and unions for each industrial sector, earnings losses studies for European workers generally find small but nevertheless statistically significant losses. Given these differences between regulated and non-regulated labor markets, it is particularly interesting to study the effects in Switzerland. This is because Swiss labor market regulations are more like in the U.S. than in many European countries, but many other economic conditions are very close to the European model (e.g., the educational system). This unique feature of Switzerland makes it interesting to

³Administrative records of the State of Pennsylvania, 1974–1986.

study the determinants of involuntary separations, as we do in a second step during our empirical analysis.

Previous literature explained the origin of earnings losses with human capital theory (Fallick, 1996). According to human capital theory (Becker, 1962), wage losses occur if separated workers have not only general human capital but also firm-specific human capital. If an individual with a large portion of specific human capital loses his or her job, the consequences of the separation will be more severe than for an individual who has more general human capital. Because firm-specific skills do not increase the workers' productivity outside a particular firm, workers with a large stock of firm-specific human capital cannot transfer to a firm in which productivity—and thus wages—will be as high. If we take this theoretical argument literally, we should always expect a loss associated with a separation, independent of the type of separation. However, such is not always the case, as many job changes—especially voluntary ones—may result in wage gains.

Somewhat different patterns can be expected when referring to Lazear's skill-weights approach (Lazear, 2009) to explain differences in earnings losses across different types of separation. In Lazear's model, all skills are general but firms use them in different combinations with different weights attached to them. The advantage of his view of human capital is that it provides a more differentiated explanation of wage changes following a separation. According to Lazear's model, the expected wage change is likely to be negative whenever turnover is involuntary, because if workers choose not to move they have difficulties in finding a firm with an adequate skill-weights profile. He argues that the presence of these difficulties implies a negative wage change in terms of expected wages (keeping market thickness constant).

Conversely, we should expect a positive wage change for voluntary leavers, because the quitters would be those who find an outside offer from a firm with a better skill-weights profile. We also investigate and discuss these issues by estimating wage changes for different types of separation in section 4.5.

4.3 Data and Descriptive Statistics

We base the analysis on data from the Swiss Labor Force Survey (SLFS) between 1996 and 2009. The Swiss Federal Statistical Office produces the SLFS annually and it is representative for the adult population permanently living in Switzerland. The main purpose of the SLFS is to provide information on the structure of the labor force and employment behavior patterns. Strict adherence to international definitions allows Swiss data to be compared with OECD, European, and U.S. data. The SLFS is a rotating panel based on a sample of some 105,000 interviews, in which four-fifths of the households from the previous year's survey are re-interviewed.

With this data set we are able to distinguish the involuntary separations from all other types of separations, in addition to the classical demographic, educational, and occupational characteristics. From a question asking the reason for the last job loss, we can isolate those who separated involuntarily from those who separated for other reasons, such as quit, retirement, injury, working conditions, limited contract, or personal issues.⁴ Thus, we are able to achieve valid estimates of the earnings pattern by observing a worker before, during, and after an involuntary separation. Since one might be concerned that asking people about the reasons for their separation might not reveal the true reasons but rather socially desirable answers, we address the problem by analyzing earnings losses for each other type of separation (e.g., quit) to see whether the respective results are consistent with what one would expect if the answers were correct. For example, we should observe that voluntary separations go together with wage gains and not wage losses.

To avoid special circumstances, such as those that might arise from retirement, our sample takes into account only individuals aged 18-65. We also restrict the sample to individuals who are either employed or unemployed, so as to avoid misspecification due to

⁴The possible categories are involuntary separation, voluntary leave, limited contract, illness or accident, forced retirement, early retirement, retirement, working conditions, education, army duty, child care, personal or family, desire for change, desire to stop working, and other.

people who are in school, retired, or not active in the labor force. The earnings variables come from the Swiss Survey on Income and Living Conditions, a very precise data source for income resulting from labor activity. Thus, for those who are unemployed at the time of the interview, we have zero earnings in the data. Among individuals who are employed, we consider only those for whom the earnings information—net annual income and hourly wage—is not missing.

After creating the panel and removing missing values, we are left with a sample of 67,590 observations. Table 4.3.1 provides descriptive statistics. Over the period 1996-2009, the incidence of a job loss due to involuntary separation is 7.07 percent. Over the same period, the average worker works 36 hours per week, earns 34.73 Swiss Francs⁵ (CHF) per hour, and has a net annual income of CHF 66,478. Almost 70 percent of the sample has a full-time employment for the entire period of observation.

The average age in the sample is 42.2 years, and 55.6 percent of the individuals in the sample are male. Regarding education, the average individual achieved 13.4 years of schooling, divided among compulsory education (almost 10 percent), vocational education (70 percent), and academic education (20 percent). In terms of educational degrees, 9 percent of the sample hold a primary education degree, whereas 58 percent achieved a secondary level and 33 percent a tertiary level.

4.4 Methods

4.4.1 The Wage Equation

To quantify the earnings losses and their temporal pattern, we employ an approach similar to Jacobson et al. (1993) by defining a worker's earnings loss as the difference between his or her observed and expected earnings had the events that led to the job loss not

⁵Wages are inflated to the year 2010. In 2010, 1 CHF = 1 USD. In Switzerland, inflation is very low and stable over time.

Table 4.3.1: DESCRIPTIVE STATISTICS

	Mean (1)	Std. Dev. (2)	Min (3)	Max (4)
Involuntary separation	0.07	0.26	0.00	1.00
Annual income	66,478	39,326	0.00	571,654
Hourly wage	34.73	16.83	0.00	173.41
Weekly working hours	35.96	12.51	0.00	97.00
Male	0.56	0.50	0.00	1.00
Age	42.22	9.84	18.00	65.00
Full-time worker	0.68	0.47	0.00	1.00
Swiss	0.75	0.44	0.00	1.00
Tenure (in years)	9.86	8.88	0.00	48.41
Years of schooling	13.39	2.44	9.00	18.00
Primary education	0.09	0.28	0.00	1.00
Secondary education	0.58	0.49	0.00	1.00
Tertiary education	0.33	0.47	0.00	1.00
Mandatory education	0.08	0.27	0.00	1.00
Vocational education	0.72	0.45	0.00	1.00
Academic education	0.20	0.40	0.00	1.00
<i>N</i>	67,590			

Notes: Swiss Labor Force Survey, Authors' calculations.

occurred. Letting y_{it} denote the earnings⁶ of worker i in period t and letting $S_{i,s} = 1$ if worker i experienced an involuntary separation at time s (and $S_{i,s} = 0$ otherwise⁷), the definition of the loss is hereafter illustrated:

$$E(y_{it} \mid S_{i,s} = 1, I_{i,s-v}) - E(y_{it} \mid S_{i,s} = 0 \ \forall \ s, I_{i,s-v}) \quad (4.4.1)$$

where $I_{i,s-v}$ is an information set at time $(s - v)$ containing individual-specific earnings determinants (either observable or unobservable), regardless of whether the individual experienced an involuntary separation. We also assume v to be sufficiently large that the events that led to separation have not yet begun.

The earnings at a given date are assumed to depend on the event of an involuntary

⁶We use the net annual income and the hourly wage as earning measures.

⁷“Otherwise” denotes individuals who did not experience an involuntary separation, but it does not necessarily exclude the possibility that they separated for other reasons. We also performed our investigation by comparing those who involuntarily separated with those who did not separate at all, with results similar to those we present in section 4.5. We stay with our approach because it better represents the typical working life of an employee.

separation and on some controls for fixed- and time-varying characteristics. Thus we can rewrite the earnings equation as follows:⁸

$$\ln(y_{it}) = \alpha_i + \gamma_t + X'_{it}\beta + \sum_{k=-4}^4 \delta_k \cdot S_{it}^k + \varepsilon_{it} \quad (4.4.2)$$

In equation (4.4.2), the individual fixed effect α_i captures the impact of time-invariant differences among individuals in observed and unobserved characteristics, and γ_t is a set of dummy variables for each year in the sample that gauges the general time pattern of earnings. The vector X_{it} consists of the observed, time-varying characteristics of the individual. As most available variables—such as experience, tenure, or industry—might be endogenous to the involuntary job loss and thus constitute a form of separation costs themselves, we restrict our controls to age, age squared, and interactions among these controls and gender.

The set of dummy variables S_{it}^k represents the event of involuntary separation and δ_k measures the effect of such job loss in the years before, during, and after separation. Specifically, $S_{it}^k = 1$ if worker i experienced an involuntary separation k years prior to (k is then negative), during (k equals zero), or since (k is positive) year t ; and $S_{it}^k = 0$ if individual i experienced no involuntary job loss during period t . Therefore, decomposing the sum $\sum \delta_k \cdot S_{it}^k$ yields a measure of the earnings loss during each year k .

4.4.2 Estimation Strategy

The first estimation strategy we use is the typical approach of the earning losses literature, i.e. estimating equation (4.4.1) by fixed-effects methods (Couch, 2001; Jacobson et al., 1993; Kletzer and Fairlie, 2003; Stevens, 1997; White, 2010; Zwick, 2012). OLS estimates with no individual fixed effects are probably biased because of the endogeneity of different types of job losses and because of the unobservable worker characteristics that may cause

⁸Consistent with the existing empirical literature, we assume the relation between our dependent variable and the explanatory variables to be logarithmic.

them. Researchers' inability to distinguish between workers with high or low productivity, leads to a sample of involuntary separations that might not properly represent the overall working population. We use the panel structure of our data to mitigate this selection bias by filtering out the time-constant heterogeneity at the individual level with the within transformation.

Two general problems, still not properly covered in the existing literature, remain. The first problem concerns the practice of interpreting the parameters of log-linear models estimated by least squares as semi-elasticity. This practice can be misleading, especially in the presence of heteroskedasticity and measurement error. With heteroskedastic errors, the log-linear model does not consistently estimate the semi-elasticities.⁹ Because of this issue, Santos Silva and Tenreyro (2006) argue that constant elasticity models should preferably not be linearized but rather estimated by Poisson pseudo maximum likelihood (PPML). Furthermore, due to Jensen's inequality, in log-linear models we cannot predict levels, because the expected value of the logarithm of a random variable is different from the logarithm of its expected value.¹⁰

The second problem is related to the specification of the dependent variable, the natural logarithm of earnings. If an individual is unemployed at the time of the interview, his or her earnings resulting from labor market activity are zero. The logarithm of zero is undefined, and we are left with an unbalanced panel. Including the zeros has at least three advantages. First, it allows us to fully exploit the sample of individuals who experienced an involuntary separation. Second, it mitigates the selection problem because job separation affects the probability of working (von Wachter et al., 2008). Third, it introduces a new measure of losses, because we otherwise exclude all spells with zero earnings. If we include the zeros, we are not only measuring the earning loss after re-employment due to a loss of specific human capital but rather the total loss resulting from an involuntary separation. We call this new measure "productivity loss," because it

⁹With heteroskedastic ε , $E(\varepsilon|x) = f(x)$ and $E(y|x) = \exp(x'\beta)f(x)$. In this case, $\partial E(y|x)/\partial x \neq \beta$.

¹⁰In case of concave functions such as $\ln(y)$, we have that $E(\ln y) \leq \ln E(y)$.

reflects the value of the forgone productivity while the worker is not active in the labor market. The unanswered question is whether the productivity loss is also long-lasting or whether it affects only the year of separation.

To simultaneously overcome both these problems, we use the PPML estimator that was introduced by Santos Silva and Tenreyro (2006) and further discussed in Santos Silva and Tenreyro (2011). The PPML estimator identifies the coefficients using the same first-order conditions that are used by the maximum-likelihood estimator derived from the Poisson distribution. However, the PPML estimator does not require the dependent variable to be Poisson distributed (Gourieroux et al., 1984). Following Cameron and Trivedi (2013), we can estimate the parameters of interest by solving the set of first-order conditions hereafter:

$$\sum_{i=1}^n [y_i - \exp(x_i\beta)] \cdot x_i = 0 \quad (4.4.3)$$

The PPML approach can be seen as a nonlinear-least-square specification of an equation with uniform weights given to observations. Without further information—or assumptions—on the pattern of heteroskedasticity, giving the same weight to all observations is the more natural way of proceeding. In terms of estimated parameters, PPML has two main differences with respect to OLS. First, as we already underlined, the coefficients of PPML do not suffer from bias that arises from estimating semi-elasticities with a log-linear model. Second, PPML allows to make predictions on wage levels and not only in terms of log-wages as in the case of OLS.

Another important feature of the PPML estimator is that it does not require the dependent variable to be an integer, and it is also consistent with over-dispersion (Fally, 2015). Santos Silva and Tenreyro (2006) and Santos Silva and Tenreyro (2011) provide additional evidence on the good performance of PPML by presenting a simulation study using different heteroskedasticity patterns, and showing that the PPML estimator is the least biased of various OLS functional forms, non-linear least squares, Tobit models, gamma pseudo-maximum-likelihood methods, and in the presence of a large fraction of

zeros in the dependent variable.

Furthermore, the PPML approach can be consistently adapted to panel data, with the use of PPML with fixed effects (Fally, 2015; Wooldridge, 1999). This PPML panel estimator is consistent to arbitrary patterns of serial correlation but requires the appropriate robust standard errors suggested by Wooldridge (1999) for valid inference. Considering all these advantages, we decided to use the PPML approach, which has never been applied to research on separations and earning losses.

4.5 Empirical Findings

4.5.1 Effect of Involuntary Separation on Annual Income and Hourly Wage

Table 4.5.1 provides regression outputs of the fixed-effects model.¹¹ The first two columns show the results with the natural logarithm of annual income as dependent variable for the entire sample (column 1) and for full-time workers only (column 2). Usually, the existing literature either makes no distinction between full-time and part-time or just focuses on full-time workers. In this chapter, we distinguish full-time workers from the entire sample to see whether those workers who go from a full-time job to another full-time job are more, less, or equally damaged by an involuntary separation.

In terms of annual income (table 4.5.1, columns 1-2), the coefficients of the post-separation variables are negative and significant at the 1 percent level, from the year of involuntary job loss up to four years thereafter. Individuals who experience an involuntary separation have—on average—an immediate loss in annual income of 10.4 percent and a long-term loss of 11.4 percent three years after separation and 11.3 percent four years after separation.¹² In terms of Swiss Francs, the typical involuntarily separated worker

¹¹Appendix Table 4.7.1 presents regression outputs using the PPML approach excluding zeros, with results in line with the one depicted in Table 4.5.1.

¹²Note that the exact effect, calculated as $\exp(\beta) - 1$, is slightly less. The exact immediate loss is 9.9

Table 4.5.1: EARNING EFFECTS OF INVOLUNTARY SEPARATIONS, FE ESTIMATION

Variables	Annual income <i>Entire sample</i> (1)	Annual income <i>Full-time workers</i> (2)	Hourly wage <i>Entire sample</i> (3)	Hourly wage <i>Full-time workers</i> (4)
3 years before separation	-0.011 (0.028)	-0.009 (0.017)	-0.019 (0.024)	-0.003 (0.017)
2 years before separation	-0.026 (0.026)	-0.028* (0.013)	-0.023 (0.022)	-0.017 (0.015)
1 year before separation	0.052 [†] (0.027)	-0.043** (0.015)	-0.035 [†] (0.021)	-0.037* (0.016)
Year of separation	-0.104** (0.032)	-0.093** (0.019)	-0.071** (0.024)	-0.091** (0.021)
1 year after separation	-0.120** (0.033)	-0.106** (0.020)	-0.084** (0.026)	-0.103** (0.021)
2 years after separation	-0.132** (0.034)	-0.104** (0.020)	-0.098** (0.026)	-0.106** (0.021)
3 years after separation	-0.114** (0.036)	-0.085** (0.024)	-0.101** (0.029)	-0.098** (0.025)
4 years after separation	-0.113** (0.039)	-0.099** (0.025)	-0.093** (0.030)	-0.106** (0.027)
Individual fixed effects	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES
R ²	0.227	0.088	0.064	0.075
N	66,284	45,250	66,284	45,250

Notes: ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.10$. Robust standard errors clustered at the individual level are in parentheses. All models include a constant and interactions between gender and age. The dependent variable is expressed in logarithm.

Swiss Labor Force Survey, Authors' calculations.

loses on average almost CHF 7,000 in the year of separation and slightly more than CHF 7,500 four years after separation, compared with their expected income had the separation never happened. Altogether, the average post-separation loss within the first four years amounts to CHF 38,760, which represents roughly 60% of a year's wage (on average).

Given that the hourly wage is a more precise proxy of employees' human capital use, columns 3 and 4 of Table 4.5.1 present estimates with the natural logarithm of hourly wage as the dependent variable. We do this additional analysis for two reasons: First, the hourly wage takes into account the potential reduction in working hours in the new job; and, second, the hourly wage filters out company bonuses and other potential biases not related to productivity. The post-separation coefficients are all negative and significant at the 1 percent level, with an immediate loss of 7.1 percent for the entire sample and 9.1 percent for full-time workers. At the same time, the loss four years after is 9.3 percent for the entire sample and 10.6 percent for full-time workers. In terms of hourly wage, workers going from a full-time employment to another full-time employment appear to suffer from larger losses, especially in the long-term. However, in general we do not find large differences between the entire sample and the sub-sample of workers going from a full-time occupation to another full-time occupation after a job loss.

As commonly found in the earning losses literature, involuntarily separated workers already experience a small earning loss one to two years before separation. According to our estimates of Table 4.5.1, this loss is on the order of 3-4 percent, and it is usually attributed to wage renegotiation aimed at avoiding separation or to negative time-varying unobservables for those who lose their jobs later (Zwick, 2012). The pre-separation losses are relatively small, and marginally but nevertheless significant.

These first estimates constitute an important result, because they give clear evidence of the consequences of an involuntary separation, an event that jeopardizes an individual's earnings not only at the moment of separation but also in the long run. This supports

percent, and the exact long-term loss four years after separation is 10.7 percent.

Table 4.5.2: EARNING EFFECTS OF INVOLUNTARY SEPARATIONS, PPML ESTIMATION

Variables	Annual income <i>Entire sample</i> (1)	Annual income <i>Full-time workers</i> (2)	Hourly wage <i>Entire sample</i> (3)	Hourly wage <i>Full-time workers</i> (4)
3 years before separation	0.004 (0.026)	0.004 (0.028)	-0.020 (0.039)	0.007 (0.028)
2 years before separation	-0.018 (0.024)	-0.023 (0.026)	-0.028 (0.038)	-0.005 (0.028)
1 year before separation	-0.069* (0.027)	-0.080** (0.030)	-0.076* (0.038)	-0.074* (0.031)
Year of separation	-0.533** (0.039)	-0.557** (0.045)	-0.531** (0.049)	-0.560** (0.047)
1 year after separation	-0.235** (0.034)	-0.255** (0.039)	-0.209** (0.045)	-0.247** (0.041)
2 years after separation	-0.204** (0.034)	-0.222** (0.038)	-0.204** (0.045)	-0.223** (0.040)
3 years after separation	-0.183** (0.039)	-0.196** (0.043)	-0.200** (0.048)	-0.202** (0.046)
4 years after separation	-0.211** (0.045)	-0.231** (0.051)	-0.206** (0.053)	-0.227** (0.053)
Individual fixed effects	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES
(Pseudo) R ²	0.220	0.211	0.082	0.092
N	67,590	45,839	67,590	45,839

Notes: ** $p < 0.01$, * $p < 0.05$, † $p < 0.10$. Robust standard errors clustered at the individual level are in parentheses. All models include a constant and interactions between gender and age. The pseudo R² is computed as the square of the correlation between the dependent variable and its fitted values.

Swiss Labor Force Survey, Authors' calculations.

the theoretical argument that involuntary separations involve a permanent loss of parts of human capital. Looking at the time pattern of earning losses, we note that there is no sign of recovery in the first four years after an involuntary separation. This is somewhat different from what the literature usually finds (Couch and Placzek, 2010), but we consider a time interval that might be too short to observe a wage recovery;¹³ studies with similar time intervals also estimate no post-separation recovery (Kletzer and Fairlie, 2003; Ruhm, 1991).

Table 4.5.2 reports the results of our second econometric approach, the PPML. As we

¹³In the displacement literature, this time interval ranges between two years after separation (Couch and Placzek, 2010) and twenty years after separation (von Wachter et al., 2008).

discussed previously, the PPML allows us to analyze workers with zero earnings from labor market activity, and thus estimate the total productivity loss caused by an involuntary separation. As for Table 4.5.1, the first two columns of Table 4.5.2 present results with the annual income as dependent variable (column 1 for the entire sample, column 2 for full-time workers), whereas columns 3 and 4 present results with the hourly wage as dependent variable (column 3 for the entire sample, column 4 for full-time workers).

A look at the set of dummies representing the post-separation productivity loss reveals that the coefficients are all negative and significant at the one percent level. In terms of annual income, involuntarily separated workers suffer from an immediate loss of 41.3 percent in the year of separation and a long-term loss of 16.7 percent and 19.0 percent in the third and fourth year after separation, respectively.¹⁴ Estimated losses for full-time workers are very close to those of the entire sample, ranging between 42.7 percent in the year of separation and 20.6 percent in the fourth year after. In terms of hourly wage the estimated losses are also similar. We find an immediate loss of 41.2 percent (per hour) and a long-term loss of 18.6 percent (per hour) for the entire sample, and an immediate loss of 42.9 percent (per hour) and a long-term loss of about 20 percent (per hour) for full-time workers. Throughout Table 4.5.2 we also estimate a statistically-significant pre-separation loss of about 6-7 percent, but the pre-separation loss is present only one year before separation.

From a theoretical point of view, the results presented in Table 4.5.1 and Table 4.5.2 suggest that workers lose at least some parts of human capital when they are fired. Given that the earning losses appear to be permanent, this also indicates that the human capital lost is not recovered in the years following an involuntary separation. This pattern is consistent with both human capital theory and Lazear's skill-weights approach. While human capital theory predicts that the losses are due to a loss of firm-specific human capital, the skill-weights approach predicts that displaced workers have difficulties to find a new firm that matches their skill profile. The two theories differ in their predictions

¹⁴Computed as $\exp(\beta) - 1$.

when we distinguish between types of separation. Human capital theory, if taken literally, implies that any type of separation is associated with an earning loss. In contrast, the skill-weights approach provides a more differentiated explanation of wage changes following a separation by clearly distinguishing between voluntary mobility (quit) and involuntary mobility (layoff). Lazear's model predicts a negative wage change whenever turnover is involuntary, because if workers choose not to move they have difficulties in finding a firm with exactly the same skill-weights profile on the external labor market. Conversely, the model predicts a positive wage change for voluntary leavers, because such workers would be those who only go if they find an outside offer with a favorable skill-weights profile. We investigate which mechanism is supported by the data in the next subsection, by analyzing wage changes after several types of separation (other than involuntary job loss).

4.5.2 Plausibility Checks

To find out whether involuntary job losses are the only type of separation that implies a significant loss, and to check for the plausibility of our self-reported separation measure, we implement the same econometric specifications to other reasons for separation. Doing so constitutes an important falsification test for the reliability of our method of identifying an involuntary separation, because we expect distinct differences in earning losses between involuntary separations and other reasons. The rationale is that involuntary job losses should be the only reason for a persistent earning loss because the other separation motives do not imply a depreciation of human capital, as suggested by Lazear (2009). To verify this hypothesis, we consider three additional motivations: separation due to (bad) working conditions, separation due to personal or family reasons, and voluntary leave.

Table 4.5.3 and Table 4.5.4 present regression results of equation (4.4.1) divided by reason for separation, estimated by fixed effects and fixed-effects PPML, respectively. For the ease of comparison, we report regression outputs for involuntary separations in

Table 4.5.3: WAGE EFFECTS FOR DIFFERENT TYPES OF JOB LOSS, FE ESTIMATION

Variables	Hourly wage <i>Involuntary separation</i> (1)	Hourly wage <i>Working conditions</i> (2)	Hourly wage <i>Personal or family reasons</i> (3)	Hourly wage <i>Voluntary separation</i> (4)
3 years before separation	-0.019 (0.024)	-0.019 (0.018)	-0.012 (0.021)	0.007 (0.018)
2 years before separation	-0.023 (0.022)	-0.019 (0.017)	-0.020 (0.023)	-0.016 (0.019)
1 year before separation	-0.035 [†] (0.021)	-0.039* (0.017)	-0.014 (0.021)	0.005 (0.018)
Year of separation	-0.071** (0.024)	-0.013 (0.018)	-0.039 [†] (0.022)	0.038* (0.020)
1 year after separation	-0.084** (0.026)	-0.005 (0.019)	-0.020 (0.023)	0.051* (0.020)
2 years after separation	-0.098** (0.026)	-0.005 (0.020)	-0.009 (0.024)	0.058** (0.021)
3 years after separation	-0.101** (0.029)	-0.003 (0.021)	-0.042 (0.026)	0.066** (0.022)
4 years after separation	-0.093** (0.030)	-0.001 (0.023)	-0.015 (0.027)	0.052* (0.024)
Individual fixed effects	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES
R ²	0.064	0.062	0.061	0.062
N	66,284	66,284	66,284	66,284

Notes: ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.10$. Robust standard errors clustered at the individual level are in parentheses. All models include a constant and interactions between gender and age. The dependent variable is expressed in logarithm.

Swiss Labor Force Survey, Authors' calculations.

each first column of both tables and we present results only for hourly wage losses.¹⁵ The first important result throughout tables 4.5.3 and 4.5.4 is that no separation reason other than involuntary job loss implies a permanent wage loss (using hourly wage or annual income as dependent variable does not change the estimated effects significantly). The estimated effects are also qualitatively the same either using fixed-effects estimation or fixed-effects PPML.

In detail, we find that separation due to (bad) working conditions implies hardly any significant loss either before or after the separation happened. The only significant

¹⁵Appendix tables 4.7.2 and 4.7.3 present regression outputs for annual income losses, but the results are qualitatively similar to those of tables 4.5.3 and 4.5.4.

Table 4.5.4: WAGE EFFECTS FOR DIFFERENT TYPES OF JOB LOSS, PPML ESTIMATION

Variables	Hourly wage <i>Involuntary separation</i> (1)	Hourly wage <i>Working conditions</i> (2)	Hourly wage <i>Personal or family reasons</i> (3)	Hourly wage <i>Voluntary separation</i> (4)
3 years before separation	-0.020 (0.039)	0.009 (0.026)	-0.000 (0.026)	0.012 (0.020)
2 years before separation	-0.028 (0.038)	-0.008 (0.023)	-0.006 (0.029)	-0.025 (0.020)
1 year before separation	-0.076* (0.038)	-0.038 (0.024)	-0.019 (0.028)	-0.006 (0.019)
Year of separation	-0.531** (0.049)	-0.042 (0.026)	-0.137** (0.029)	0.016 (0.021)
1 year after separation	-0.209** (0.045)	0.007 (0.026)	-0.027 (0.030)	0.043* (0.022)
2 years after separation	-0.204** (0.045)	-0.005 (0.027)	0.002 (0.031)	0.046* (0.023)
3 years after separation	-0.200** (0.048)	-0.006 (0.028)	-0.035 (0.033)	0.058* (0.025)
4 years after separation	-0.206** (0.053)	-0.003 (0.032)	-0.007 (0.036)	0.033 (0.028)
Individual fixed effects	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES
(Pseudo) R ²	0.082	0.066	0.067	0.067
N	67,590	67,590	67,590	67,590

Notes: ** $p < 0.01$, * $p < 0.05$, † $p < 0.10$. Robust standard errors clustered at the individual level are in parentheses. All models include a constant and interactions between gender and age. The pseudo R² is computed as the square of the correlation between the dependent variable and its fitted values. Swiss Labor Force Survey, Authors' calculations.

loss is one year before separation, a loss with a magnitude of less than 4 percent (table 4.5.3, column 2). Workers separated because of family or personal reasons experience a loss only in the year of separation. Their loss is rather small, ranging between 4 percent using fixed-effect estimation and about 13 percent using PPML. In terms of significance, in the fixed-effect model the loss is marginally significant, whereas in the PPML model the loss is significant at the 1 percent level.

The last type of separation we consider is voluntary separation or quits (column 4 of tables 4.5.3 and 4.5.4). For workers who voluntarily quit their jobs we observe a statistically significant increase in wage in the years following the separation. This post-separation increase is permanent and ranging between 3 to 6 percent depending on the estimation method applied. This is as expected by Lazear's skill-weights approach. According to Lazear, while the expected wage loss is necessarily positive if turnover is involuntary, we should expect a wage gain for voluntary leavers, because the quitters would be those who find an outside offer from a firm with a better skill-weight profile. Therefore, we conclude that our way of identifying the type of separation produces valid results.

4.5.3 The Determinants of Involuntary Separations

In the presence of such relevant losses following an involuntary separation, one corollary question arises: What determines an involuntary separation? Specifically, we want to investigate whether education level and education type are associated with a lower probability of involuntary separation. On the relation between education and involuntary separation, it has been previously shown that education has monetary returns such as reduced unemployment risk (Mincer, 1991), shorter unemployment spells (Kettunen, 1997), and smaller earning losses following a displacement (Eliason and Storrie, 2006; White, 2010). More generally, Farber (2010) finds that while the job loss rate of more educated workers increased during the period 1984–2002, less educated ones continue to have the

highest rates of job loss overall. We complement these results by analyzing whether this is also valid for involuntary separations and whether the relation holds for all educational types.

To fill this gap we estimate a Probit model¹⁶ in which the dependent variable equals one if an individual suffered from involuntary separation during the period 1996–2009 and zero otherwise.¹⁷ According to the official definitions of the Swiss State Secretariat for Education and Research, we code education in three ways according to the highest degree obtained: years of schooling, education level (primary, secondary, and tertiary), and type of education¹⁸ (mandatory, vocational, and academic). Beyond the educational variables, we include demographic characteristics, region fixed effects, time fixed effects, and industry fixed effects according to the General Classification of Economic Activities (NOGA).

Table 4.5.5 presents the results. Focusing on the explanatory variables related to education, we estimate an average marginal effect of -0.4 percentage points for a one-unit increase in years of schooling (column 1). In terms of predicted probabilities, the overall likelihood of experiencing an involuntary separation is 7.07 percent (either estimated with the Probit model or just by looking at descriptive statistics). An individual with 9 years of schooling (compulsory education) has a predicted probability of experiencing an involuntary job loss of 11.0 percent, which is significantly higher than the sample average. An individual with 13 years of schooling (high school degree or vocational maturity) has a 6.2-percent predicted probability of being dismissed, which is below the sample average and almost half the probability of those holding a compulsory education as their highest degree. For individuals who completed a university degree or a higher vocational degree (18 years of education and training), the predicted probability of experiencing an

¹⁶Appendix Table 4.7.4 presents regression outputs for linear probability models, with almost no difference from the Probit results.

¹⁷Here we do not claim the estimated effects to be causal, because we suspect the education variables to be endogenous in the job loss equation. We mitigate this problem by including many control variables.

¹⁸Due to data limitations, we have to ignore individuals with mixed educational paths, i.e., individuals with both academic and vocational degrees. According to Tuor and Backes-Gellner (2010), the proportion of individuals with a mixed path in Switzerland are roughly 10 percent of the working population.

Table 4.5.5: DETERMINANTS OF INVOLUNTARY SEPARATION, PROBIT MODELS

Variables	Involuntary separation dY/dX (1)	Involuntary separation dY/dX (2)	Involuntary separation dY/dX (3)
Age	0.002 (0.002)	0.002 (0.002)	0.002 (0.002)
Age squared/100	-0.001 (0.002)	-0.001 (0.002)	-0.000 (0.002)
Tenure (in years)	-0.018** (0.001)	-0.018** (0.001)	-0.018** (0.001)
Tenure squared/100	0.036** (0.002)	0.036** (0.002)	0.036** (0.002)
Swiss	-0.021** (0.005)	-0.021** (0.005)	-0.024** (0.005)
Male	0.007 (0.005)	0.005 (0.005)	0.003 (0.004)
Years of schooling	-0.004** (0.001)		
<i>Level of education</i>			
Secondary education		-0.008 (0.007)	
Tertiary education		-0.025** (0.008)	
<i>Type of education</i>			
Vocational education			-0.006 (0.008)
Academic education			-0.026** (0.009)
Industry fixed effects	YES	YES	YES
Time fixed effects	YES	YES	YES
Region fixed effects	YES	YES	YES
(Pseudo) R ²	0.170	0.170	0.169
N	67,590	67,590	67,590

Notes: ** $p < 0.01$, * $p < 0.05$, † $p < 0.10$. Robust standard errors clustered at the individual level are in parentheses. All models include a constant. The pseudo R² is computed as the square of the correlation between the dependent variable and its fitted values.

Swiss Labor Force Survey, Authors' calculations.

involuntary job loss is only 5.9 percent. Thus, the more years of education, the better a worker is protected against involuntary job losses.

We obtain similar results and predicted probabilities in column 2, where we split education into primary (base category), secondary, and tertiary levels. According to our estimates, having a tertiary degree—either vocational or academic—is the best protection against involuntary job loss, reducing the overall probability of separation by one third.

Column 3 focuses on the type of education, divided into mandatory (base category), vocational, and academic. Both vocational and academic tracks are negatively correlated with the probability of an involuntary separation, but only the coefficient on academic education is highly significant. Compared to those holding a mandatory education, having an academic degree has a negative 2.6-percentage-point effect on the job loss probability. Compared to workers with a vocational degree, workers having an academic degree are less likely to experience an involuntary separation by 2 percentage points. In terms of predicted probabilities, having mandatory education as highest level is associated with a job loss probability of 10.2 percent, while having a vocational or academic degree reduces the job loss probability down to 6.9 percent and 6.1 percent, respectively.

4.6 Conclusions

Using Swiss Labor Force Survey data from 1996 to 2009, we estimate the earning losses of workers experiencing an involuntary job loss. We follow two empirical approaches: the ordinary fixed-effect method and the fixed-effects PPML approach, a method new in the literature of job losses that allows considering the full set of involuntary separations, even those workers with zero earnings because of unemployment. Using the first approach, we estimate an immediate loss of about 10 percent, a loss remaining statistically significant four years after separation at about 11 percent. Using the second method, we estimate a loss of 40 percent in the year of separation and a long-term loss of about 20 percent

four years after separation. This second estimates are larger, because we are including individuals with zero earnings from labor market activity, thus taking into consideration that their labor productivity is fully lost during that time. We term this second loss “total productivity loss,” because it captures the total losses that occur after an involuntary separation and that represent the productivity that is foregone due to the job loss and loss of firm-specific productivity.

We also find that while involuntary separations cause permanent scars, the other types of separation—as expected—cause either light blemishes or even wage gains, as in the case of voluntary leave. These results compelled us to study whether education level or type is related with a lower probability of involuntary separation. We complement the existing literature and find that tertiary education—either academic or vocational—plays a major role in reducing the risk of job loss.

We are the first study to use fixed-effects PPML to estimate the earning losses after a job separation, and we contribute to the methodological literature by adding a new approach to analyze separation-related losses. The fixed-effects PPML has at least three advantages: First of all, using the PPML approach we can identify the total productivity loss following an involuntary separation, and not only the post-separation earning loss of workers after they found a new job (we also account for the time to find a new job and the earnings that are forgone during that period). Second, it is more appropriate when the relationship between the dependent variable and the covariates is logarithmic and it performs very well when the dependent variable is non-negative. Third, it allows including individuals that have zero earnings from labor market activity with relatively soft assumptions and without manipulating the dependent variable.

Our study could be extended in a number of ways. For example, the event of an involuntary job loss might not be completely (conditionally) exogenous, which could bias our estimated effects. We are especially worried about time-varying unobservable characteristics that affect a worker’s performance and thus impact the likelihood of getting

dismissed. Unfortunately, it is difficult to find appropriate exclusion restrictions in the job loss literature, and unless relying on strong assumptions and use the matrix of (internal) instruments suggested by Hausman and Taylor (1981), we cannot use instrumental variable approaches to estimate causal effects. Using matching estimators is also not very helpful in our case, because we would match individuals according to observable characteristics, while our true interest would lie in individuals' unobserved traits—or at least valid proxies for such traits. We thus prefer to mitigate the potential endogeneity by filtering out the time-invariant heterogeneity among individuals and assuming the impact of bad time-varying unobservables to be ignorable.

To conclude, in this chapter we were interested in the window of risks and/or opportunities that individuals face after an involuntary separation. Our estimates suggest that there is a window of opportunity that does indeed open following a job separation, because most of the sample finds a job after a separation. However, those who manage to climb through will forsake their original earning projections for at least the following four years. This result highlights the importance of reducing the likelihood of an involuntary job loss before it happens, given the near impossibility of avoiding persistent losses once the separation occurs. A better academic as well as a better vocational education help to prevent such involuntary job losses, as demonstrated by the findings here and elsewhere.

4.7 Appendix

4.7.1 PPML without Zeros

Table 4.7.1: EARNING EFFECTS OF INVOLUNTARY SEPARATIONS, PPML WITHOUT ZEROS

Variables	Annual income <i>Entire sample</i> (1)	Annual income <i>Full-time workers</i> (2)	Hourly wage <i>Entire sample</i> (3)	Hourly wage <i>Full-time workers</i> (4)
3 years before separation	0.064 (0.018)	0.004 (0.016)	-0.017 (0.034)	0.007 (0.017)
2 years before separation	-0.017 (0.016)	-0.026* (0.013)	-0.024 (0.033)	-0.012 (0.016)
1 year before separation	-0.033 [†] (0.018)	-0.035* (0.015)	-0.037 (0.032)	-0.029 [†] (0.017)
Year of separation	-0.091** (0.021)	-0.080** (0.018)	-0.069* (0.034)	-0.078** (0.023)
1 year after separation	-0.104** (0.022)	-0.087** (0.019)	-0.073* (0.035)	-0.079** (0.023)
2 years after separation	-0.110** (0.022)	-0.100** (0.020)	-0.102** (0.035)	-0.100** (0.024)
3 years after separation	-0.085** (0.026)	-0.079** (0.025)	-0.095* (0.037)	-0.085** (0.028)
4 years after separation	-0.101** (0.027)	-0.104** (0.025)	-0.092* (0.041)	-0.097** (0.032)
Individual fixed effects	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES
(Pseudo) R ²	0.210	0.198	0.059	0.079
N	66,255	44,884	66,255	44,884

Notes: ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.10$. Robust standard errors clustered at the individual level are in parentheses. All models include a constant and interactions between gender and age. The pseudo R² is computed as the square of the correlation between the dependent variable and its fitted values.

Swiss Labor Force Survey, Authors' calculations.

4.7.2 Additional Robustness Checks

Table 4.7.2: INCOME EFFECTS FOR DIFFERENT TYPES OF JOB LOSS, FE ESTIMATION

Variables	Annual Income <i>Involuntary separation</i> (1)	Annual Income <i>Working conditions</i> (2)	Annual Income <i>Personal or family reasons</i> (3)	Annual Income <i>Voluntary separation</i> (4)
3 years before separation	-0.011 (0.028)	0.000 (0.024)	-0.026 (0.025)	0.033 (0.022)
2 years before separation	-0.027 (0.026)	-0.007 (0.023)	-0.056* (0.026)	0.005 (0.025)
1 year before separation	-0.052 [†] (0.027)	-0.046 [†] (0.026)	-0.054* (0.024)	0.020 (0.026)
Year of separation	-0.104** (0.032)	-0.009 (0.027)	-0.058* (0.027)	0.098** (0.027)
1 year after separation	-0.120** (0.033)	-0.002 (0.028)	-0.050 [†] (0.028)	0.108** (0.028)
2 years after separation	-0.132** (0.034)	0.006 (0.029)	-0.019 (0.030)	0.114** (0.029)
3 years after separation	-0.114** (0.036)	0.004 (0.031)	-0.040 (0.031)	0.120** (0.030)
4 years after separation	-0.113** (0.039)	0.003 (0.033)	-0.036 (0.034)	0.117** (0.033)
Individual fixed effects	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES
R ²	0.227	0.227	0.228	0.228
N	66,284	66,284	66,284	66,284

Notes: ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.10$. Robust standard errors clustered at the individual level are in parentheses. All models include a constant and interactions between gender and age. The dependent variable is expressed in logarithm.

Swiss Labor Force Survey, Authors' calculations.

Table 4.7.3: INCOME EFFECTS FOR DIFFERENT TYPES OF JOB LOSS, PPML ESTIMATION

Variables	Annual Income <i>Involuntary separation</i> (1)	Annual Income <i>Working conditions</i> (2)	Annual Income <i>Personal or family reasons</i> (3)	Annual Income <i>Voluntary separation</i> (4)
3 years before separation	0.004 (0.026)	-0.007 (0.022)	-0.008 (0.026)	0.019 (0.019)
2 years before separation	-0.018 (0.024)	-0.010 (0.021)	-0.033 (0.029)	-0.024 (0.020)
1 year before separation	-0.069* (0.027)	-0.053* (0.022)	-0.045 [†] (0.027)	-0.004 (0.021)
Year of separation	-0.533** (0.039)	-0.058* (0.023)	-0.147** (0.029)	0.029 (0.021)
1 year after separation	-0.235** (0.034)	-0.015 (0.024)	-0.060* (0.029)	0.050* (0.021)
2 years after separation	-0.204** (0.034)	-0.025 (0.025)	-0.012 (0.031)	0.055* (0.023)
3 years after separation	-0.183** (0.039)	-0.025 (0.026)	-0.029 (0.032)	0.064** (0.024)
4 years after separation	-0.211** (0.045)	-0.021 (0.031)	-0.024 (0.033)	0.055* (0.027)
Individual fixed effects	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES
(Pseudo) R ²	0.220	0.209	0.201	0.210
N	67,590	67,590	67,590	67,590

Notes: ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.10$. Robust standard errors clustered at the individual level are in parentheses. All models include a constant and interactions between gender and age. The pseudo R² is computed as the square of the correlation between the dependent variable and its fitted values.

Swiss Labor Force Survey, Authors' calculations.

4.7.3 Linear Probability Models

Table 4.7.4: DETERMINANTS OF INVOLUNTARY SEPARATION, LPM

Variables	Involuntary separation <i>Dummy=1 if Yes</i> (1)	Involuntary separation <i>Dummy=1 if Yes</i> (2)	Involuntary separation <i>Dummy=1 if Yes</i> (3)
Age	0.006** (0.002)	0.006** (0.002)	0.005** (0.002)
Age squared/100	-0.004 [†] (0.002)	-0.004 [†] (0.002)	-0.003 (0.002)
Tenure (in years)	-0.018** (0.001)	-0.018** (0.001)	-0.018** (0.001)
Tenure squared/100	0.039** (0.002)	0.039** (0.002)	0.039** (0.002)
Swiss	-0.023** (0.006)	-0.023** (0.006)	-0.026** (0.006)
Male	0.010* (0.005)	0.008 [†] (0.005)	0.005 (0.005)
Years of schooling	-0.005** (0.001)		
<i>Level of education</i>			
Secondary education		-0.008 (0.009)	
Tertiary education		-0.026** (0.010)	
<i>Type of education</i>			
Vocational education			-0.005 (0.010)
Academic education			-0.025* (0.011)
Industry fixed effects	YES	YES	YES
Time fixed effects	YES	YES	YES
Region fixed effects	YES	YES	YES
R ²	0.082	0.081	0.081
N	67,590	67,590	67,590

Notes: ** $p < 0.01$, * $p < 0.05$, [†] $p < 0.10$. Robust standard errors clustered at the individual level are in parentheses.

All models include a constant.

Swiss Labor Force Survey, Authors' calculations.

Chapter 5

Final Remarks

A well-educated population is essential for both economic and social development. It is mainly for this reason that societies have a real interest in ensuring that children and adults have access to a wide range of educational opportunities. Although researchers and policymakers indeed acknowledge the central role of education in everyone's life, there is ambiguity about causal relationships and heterogeneous effects.

The aim of this dissertation was providing a detailed analysis of the heterogeneous effects of investments in education and educational practices at different stages of an individual's life. The idea is that averages alone provide an incomplete picture about who gains more from a given educational practice or investment in education. The first study exploits experimental data to investigate the distributional effects of smaller classes and classes with a teacher's aide. The results reveal at least two novel contributions. First, we showed that mid-achieving pupils are the ones that benefit the most from a small-class learning environment. In contrast, neither low achievers nor top achievers appear

to fully reap the gains of being in a small class. These findings are both encouraging and worrisome at the same time. On the one hand, creating smaller classes is an efficient strategy to increase mean achievement, because such strategy reaches—above all—the average student. On the other hand, however, creating smaller classes does not necessarily help close the achievement gap between low achievers and high achievers.

Second, we showed that attending regular classes with a teacher's aide is beneficial for low-achieving pupils. While having an additional instructor makes no difference for most of the children, the additional support has a positive and significant effect on the bottom quintile of the achievement distribution. Interestingly, such effect is even stronger for boys and pupils coming from a disadvantaged background (i.e., black children or free-lunch eligible children). These findings on teacher's aide are encouraging, especially in terms of reducing inequality. While the net effect of smaller classes on the achievement gap is not clear, I show that adding a teacher's aide would be an effective policy for attenuating the achievement gap, particularly for classes with large shares of black or low-income students.

The second study in this dissertation examined the casual impact of education over the distribution of wages. By taking into account both the endogeneity of educational attainment and the heterogeneity of schooling effects, we contribute to the literature in at least three ways. First, we provide evidence that no unique causal impact of schooling exists and that for each individual the effect is usually above or below the estimates extensively documented by the empirical literature so far. Second, the results show that the highest returns to education occur at the lower part of the wage distribution, which means that the slope of the education-wage relationship is steeper for workers with lower wages. If we assume that the distribution of wages reflects the distribution of abilities in the workforce, which is a standard assumption in labor economics, the findings suggest that less able individuals profit more from one additional year of education. In other words, although higher-ability individuals earn on average higher wages, the slope of

their wage-education profile is flatter than that for lower-ability individuals.

Third, we provide a correlational analysis of returns between and within the academic and the vocational track. While most of the empirical literature supports the idea that academic education yields the highest returns, we demonstrate that this fact does not hold over the entire distribution of wages. Above the median wage, individuals with an academic background have higher returns than individuals with a vocational background. However, at lower parts of the wage distribution, vocational education brings higher returns than academic education. These results imply that answering the question of which type of education has higher labor market returns is not as easy as it might have once appeared from descriptive statistics or mean regression.

The third study of this dissertation examined the earnings penalties faced by workers when they experience an involuntary job loss. The first relevant contribution of this chapter consists of the finding that the wage losses following an involuntary separation are significant and long-lasting. Separated workers suffer from an immediate wage loss of about ten percent, which remains significant at least up to four years after separation. Moreover, if we include individuals with zero earnings in the analysis, both the immediate and the long-run loss almost double. The second main contribution is the analysis of different reasons for job separation, and finding that only involuntary separations cause permanent scars. All other types of separation imply either light blemishes or wage gains, as in the case of voluntary leave. This result highlights the importance of reducing the risk of experiencing an involuntary separation, given the near impossibility of avoiding persistent losses once the separation occurs. According to our estimates, acquiring more years of education as well as pursuing a tertiary academic or vocational education significantly reduces the likelihood of an involuntary separation.

Taken together, the chapters of this dissertation clearly indicate at least three important policy considerations. First, in order to design *effective* educational policies, policymakers need to understand the heterogeneous effects of those policies. Either for

a primary school intervention or a labor market adjustment, it is crucial to know who is affected by such intervention and by how much. For example, this dissertation shows that for less able individuals investing in education is more profitable and that the best instructional practice for their learning is having a teacher's aide in class. Second, depending on the *target* group of a given intervention, policymakers need to understand how the intervention affects that particular group. Simply looking at average effects on the entire population is certainly useful, but not sufficient. As this dissertation shows, educational practices, returns to education, and even job loss determinants are highly heterogeneous according to individuals' observed and unobserved characteristics.

The third consideration, which brings together the first two, is probably the most relevant in economics of education: *efficiency*. Given that resources (e.g., funds, teachers, and infrastructures) are limited, public policy has to design efficient educational policies. For example, if the objective is to increase average student test scores, then reducing class size is probably the most efficient way. However, if the objective is to rise the performance of low-achieving students or close the achievement gap, then the most efficient way is introducing a teacher's aide. Therefore, due to the heterogeneous effects that almost all educational interventions imply, policymakers have to clearly define the target group of the intervention, understand how the intervention would affect that specific group, and then design the most efficient way to achieve the objective.

One additional remark concerns the generalizability of the results of this dissertation. There are two dimensions we should consider before thinking about how it is possible to generalize my findings to other contexts. The first dimension is purely spatial: what other countries can learn from results of this dissertation? The question affects especially the policy implications based upon Swiss data. Institutions differ across countries, and although Swiss labor regulations are more like in the U.S. than in most European countries, many other economic conditions are very close to the European model (e.g., the educational system). Therefore, on the one hand, Switzerland's uniqueness allows

answering research questions that cannot be resolved by looking at other countries. On the other hand, however, it might be too ambitious to entirely extend the results based on Swiss data to other countries. Still, there are some lessons to be learned from the Swiss case. For example, many Western countries undertook a compulsory education reform in the last forty years, and the approach I follow in this dissertation can be easily adapted to other countries. Thus, while the policy implications suggested in this dissertation might be limited to the Swiss domain, the research design I use can be applied to several contexts.

The second dimension we should consider is time. How can results obtained from data from the past still be useful in today's policy? This second question concerns primarily the findings based upon Project STAR, which is the oldest data source I use in the dissertation. For example, we might be concerned that early grades in the late 1980s and early 1990s are completely different for the ones of today. This concern is legitimate but evidence shows, however, that the technology used in today's education did not change much since World War II, especially for early grades (Goldin and Katz, 2009). Moreover, the use of technology in school appears to have no impact on any outcome analyzed (Bulman and Fairlie, 2015). Therefore, I am confident that the results and patterns revealed by the STAR data are still actual and informative for policymakers.

The empirical findings and their policy implications described in this dissertation point toward several avenues for future research. Regarding the effects of class size, many researchers highlight the importance of character skills (Heckman et al., 2013). One promising stream of research, initiated by Duckworth et al.,¹ emphasizes the role of perseverance and passion for long-term goal on educational achievement and labor market success. As these skills are acquired and better developed in early grades, Project STAR offers an optimal setting for studying perseverance and long-term commitment. Moreover, almost all follow-up surveys of Project STAR contain enough information to build a "grit score," and future research might explore this avenue in more detail.

¹See for example Eskreis-Winkler et al. (2014).

Regarding the impact of education on labor market outcomes, this dissertation points at several policy implications, as we discussed in the previous chapters. However, we did not investigate the evolution over time of such effects. For example, it would be interesting to examine how the distribution of returns to education evolved in the last decades. The main policy question would be whether education mitigates wage and income inequality over the years as suggested by Brunello et al. (2009). Our results would indicate that education indeed has an inequality-reducing effect over time, because individuals at the lower quantiles of the wage distribution appear to profit more from formal schooling. However, a more thorough analysis is needed to provide a definitive answer to such question.

Another potential research avenue pertains the evolution over time of earning losses after a separation. Although this dissertation unequivocally shows that earning losses after involuntary separation are large and persistent, we do not know whether this pattern has worsened in the last three decades. Related research for the U.S. suggests that both long term unemployment risk and immediate wage cuts of laid off workers increased since the 1970s (Farber, 2010). However, it is still unclear whether the persistence of the earning losses worsened over time and whether today's determinants of being laid off are the same as the determinants of thirty or forty years ago. Further analyses of the wage loss patterns after lay-off and the determinants of lay-offs would help employees, employers, and policymakers designing better contracts and social security systems in the future.

In conclusion, my dissertation shows that typical estimates of the effects of education on the average individual provide an incomplete characterization of the actual impact of education on cognitive skills and labor market outcomes, and thus constitute a weak guide for public policy. The results suggest that size and significance of such effects are heterogeneous, depending on individual observable and unobservable characteristics. Therefore, going back to Plato's opening quote, my dissertation demonstrates that education does

indeed determine our future lives, but in a different manner for everyone of us.

Bibliography

- Abadie, A., Angrist, J., and Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1):91–117.
- Acemoglu, D. and Angrist, J. (2001). How large are human-capital externalities? Evidence from compulsory-schooling laws. In *NBER Macroeconomics Annual 2000, Volume 15*, NBER Chapters, pages 9–74. National Bureau of Economic Research.
- Addison, J. T. and Portugal, P. (1989). Job displacement, relative wage changes, and duration of unemployment. *Journal of Labor Economics*, 7(3):281–302.
- Akerlof, G. A. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.
- Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014.
- Angrist, J. D. and Lavy, V. (1999). Using maimonides’ rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2):533–575.
- Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press.
- Arabsheibani, R. and Staneva, A. V. (2012). Returns to education in Russia: Where there is risky sexual behaviour here is also an instrument. IZA Discussion Paper 6726, Institute for the Study of Labor (IZA).

- Arcand, J.-L., D'Hombres, B., and Gyselinck, P. (2005). Instrument choice and the returns to education: New evidence from Vietnam. *Labor and Demography* 0510011, EconWPA.
- Arias, O., Hallock, K. F., and Sosa-Escudero, W. (2001). Individual heterogeneity in the returns to schooling: Instrumental variables quantile regression using twins data. *Empirical Economics*, 26(1):7–40.
- Ashenfelter, O. and Krueger, A. B. (1994). Estimates of the economic returns to schooling from a new sample of twins. *American Economic Review*, 84(5):1157–1173.
- Ashenfelter, O. and Rouse, C. (1998). Income, schooling, and ability: Evidence from a new sample of identical twins. *The Quarterly Journal of Economics*, 113(1):253–284.
- Atella, V., Pace, N., and Vuri, D. (2008). Are employers discriminating with respect to weight? European evidence using quantile regression. *Economics and Human Biology*, 6(3):305–329.
- Autor, D. H., Houseman, S. N., and Kerr, S. P. (2012). The effect of work first job placements on the distribution of earnings: An instrumental variable quantile regression approach. NBER Working Paper 17972, National Bureau of Economic Research.
- Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *The Journal of Political Economy*, 70(1):9–49.
- Bertrand, M. and Pan, J. (2013). The trouble with boys: social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics*, 5(1):32–64.
- Bettinger, E., Kremer, M., and Saavedra, J. E. (2010). Are educational vouchers only redistributive? *The Economic Journal*, 120(546):F204–F228.

- Betts, J. R. and Shkolnik, J. L. (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review*, 19(1):1–15.
- Bitler, M. P., Gelbach, J. B., and Hoynes, H. W. (2006). What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review*, 96(4):988–1012.
- Borah, B. J. and Basu, A. (2013). Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence. *Health Economics*, 22(9):1052–1070.
- Braga, M., Checchi, D., and Meschi, E. (2013). Educational policies in a long-run perspective. *Economic Policy*, 28(73):45–100.
- Brunello, G., Fabbri, D., and Fort, M. (2013). The causal effect of education on body mass: Evidence from europe. *Journal of Labor Economics*, 31(1):195–223.
- Brunello, G., Fort, M., and Weber, G. (2009). Changes in compulsory schooling, education and the distribution of wages in Europe. *The Economic Journal*, 119(536):516–539.
- Buchinsky, M. (1994). Changes in the us wage structure 1963-1987: Application of quantile regression. *Econometrica*, 62(2):405–58.
- Bulman, G. and Fairlie, R. W. (2015). Technology and education: The effects of computers, the internet and computer assisted instruction on educational outcomes. volume 5 of *Handbook of the Economics of Education*. Elsevier.
- Busemeyer, M. R. and Trampusch, C. (2012). *The Political Economy of Collective Skill Formation*. Oxford University Press.
- Butcher, K. F. and Case, A. (1994). The effect of sibling sex composition on women’s education and earnings. *The Quarterly Journal of Economics*, 109(3):531–563.

- Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427.
- Cameron, A. C. and Trivedi, P. K. (2013). *Regression Analysis of Count Data*. Cambridge University Press.
- Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. NBER Working Paper 4483, National Bureau of Economic Research.
- Card, D. (1994). Earnings, schooling, and ability revisited. NBER Working Paper 4832, National Bureau of Economic Research.
- Card, D. (1999). The causal effect of education on earnings. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3 of *Handbook of Labor Economics*, chapter 30, pages 1801–1863. Elsevier.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160.
- Card, D. (2012). Earnings, schooling, and ability revisited. *Research in Labor Economics*, 35:111–136.
- Card, D. and Krueger, A. B. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *The Journal of Political Economy*, 100(1):1–40.
- Carrington, W. J. (1993). Wage losses for displaced workers: Is it really the firm that matters? *Journal of Human Resources*, 28(3):435–462.
- Chen, L.-A. and Portnoy, S. (1996). Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equation models. *Communications in Statistics-Theory and Methods*, 25(5):1005–1032.

- Chernozhukov, V. and Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132(2):491–525.
- Chernozhukov, V. and Hansen, C. (2008). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, 142(1):379–398.
- Chernozhukov, V. and Hansen, C. (2013). Quantile models with endogeneity. *Annual Review of Economics*, 5(1):57–81.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, 126(4):1593–1660.
- Couch, K. A. (2001). Earnings losses and unemployment of displaced workers in Germany. *Industrial and Labor Relations Review*, 54(3):559–572.
- Couch, K. A. and Placzek, D. W. (2010). Earnings losses of displaced workers revisited. *American Economic Review*, 100(1):572–89.
- Curti, M. (1998). Does economic stagnation affect unemployed workers, even when re-employed? *International Journal of Manpower*, 19(6):410–423.
- Dearden, L., McIntosh, S., Myck, M., and Vignoles, A. (2002). The returns to academic and vocational qualifications in Britain. *Bulletin of Economic Research*, 54(3):249–274.
- Dee, T. S. (2004). Are there civic returns to education? *Journal of Public Economics*, 88(9):1697–1720.
- Dee, T. S. and West, M. R. (2011). The non-cognitive returns to class size. *Educational Evaluation and Policy Analysis*, 33(1):23–46.
- DePaola, M., Ponzio, M., and Scoppa, V. (2013). Class size effects on student achievement: Heterogeneity across abilities and fields. *Education Economics*, 21(2):135–153.

- Dickson, M. (2013). The causal effect of education on wages revisited. *Oxford Bulletin of Economics and Statistics*, 75(4):477–498.
- Dickson, M. and Harmon, C. (2011). Economic returns to education: what we know, what we don’t know, and where we are going—some brief pointers. *Economics of Education Review*, 30(6):1118–1122.
- Ding, W. and Lehrer, S. (2011). Experimental estimates of the impacts of class size on test scores: Robustness and heterogeneity. *Education Economics*, 19(3):229–252.
- Doyle, O., Harmon, C., Heckman, J. J., Logue, C., and Moon, S. (2013). Measuring investment in human capital formation: An experimental analysis of early life outcomes. NBER Working Paper 19316, National Bureau of Economic Research.
- Doyle, O., Harmon, C. P., Heckman, J. J., and Tremblay, R. E. (2009). Investing in early human development: Timing and economic efficiency. *Economics and Human Biology*, 7(1):1–6.
- Duckworth, A. L., Peterson, C., Matthews, M. D., and Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6):1087–1101.
- Eliason, M. and Storrie, D. (2006). Lasting or latent scars? Swedish evidence on the long-term effects of job displacement. *Journal of Labor Economics*, 24(4):831–856.
- Eren, O. (2009). Does membership pay off for covered workers? A distributional analysis of the free-rider problem. *Industrial and Labor Relations Review*, 62(3):367–380.
- Eskreis-Winkler, L., Duckworth, A. L., Shulman, E. P., and Beal, S. (2014). The grit effect: predicting retention in the military, the workplace, school and marriage. *Frontiers in Psychology*, 5:5–36.
- Evans, W. N. and Montgomery, E. (1994). Education and health: Where there’s smoke

- there's an instrument. NBER Working Paper 4949, National Bureau of Economic Research.
- Fallick, B. C. (1996). A review of the recent empirical literature on displaced workers. *Industrial and Labor Relations Review*, 50(1):5–16.
- Fally, T. (2015). Structural gravity and fixed effects. *Journal of International Economics*, 97(1):76–85.
- Fang, H., Eggleston, K. N., Rizzo, J. A., Rozelle, S., and Zeckhauser, R. J. (2012). The returns to education in China: Evidence from the 1986 compulsory education law. NBER Working Paper 18189, National Bureau of Economic Research.
- Farber, H. S. (2003). Job loss in the United States, 1981-2001. NBER Working Paper 9707, National Bureau of Economic Research.
- Farber, H. S. (2010). Job loss and the decline in job security in the United States. In *Labor in the New Economy*, NBER Chapters, pages 223–262. National Bureau of Economic Research.
- Fasih, T., Kingdon, G., Patrinos, H. A., Sakellariou, C., and Soderbom, M. (2012). Heterogeneous returns to education in the labor market. Policy Research Working Paper 6170, The World Bank.
- Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, 116(3):429–456.
- Finn, J. D. and Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, 27(3):557–577.
- Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 75(1):259–276.
- Firpo, S., Fortin, N. M., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3):953–973.

- Fletcher, J. M. (2009). Is identification with school the key component in the black box of education outcomes? Evidence from a randomized experiment. *Economics of Education Review*, 28(6):662–671.
- Folger, J. and Breda, C. (1989). Evidence from Project STAR about class size and student achievement. *Peabody Journal of Education*, 67(1):17–33.
- Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. volume 4 of *Handbook of Labor Economics*, chapter 1, pages 1–102. Elsevier.
- Fournier, J.-M. and Koske, I. (2013). Public employment and earnings inequality: An analysis based on conditional and unconditional quantile regressions. *Economics Letters*, 121(2):263–266.
- Frandsen, B. R. (2012). Why unions still matter: The effects of unionization on the distribution of employee earnings. *Job Market Paper, Massachusetts Institute of Technology*.
- Fredriksson, P., Öckert, B., and Oosterbeek, H. (2014). Inside the black box of class size: Mechanisms, behavioral responses, and social background. IZA Discussion Paper 8019, Institute for the Study of Labor (IZA).
- Frölich, M. (2007). Propensity score matching without conditional independence assumption with an application to the gender wage gap in the United Kingdom. *The Econometrics Journal*, 10(2):359–407.
- Frölich, M. and Melly, B. (2010). Estimation of quantile treatment effects with stata. *Stata Journal*, 10(3):423–457.
- Gibbons, R. and Katz, L. F. (1991). Layoffs and lemons. *Journal of Labor Economics*, 9(4):351–380.
- Goldin, C. D. and Katz, L. F. (2009). *The Race Between Education and Technology*. Harvard University Press.

- Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica*, 52(3):701–20.
- Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica*, 45(1):1–22.
- Griliches, Z. and Mason, W. M. (1972). Education, income, and ability. *The Journal of Political Economy*, 80(3):74–103.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (2011). *Robust Statistics: The Approach Based on Influence Functions*, volume 114. John Wiley & Sons.
- Hanushek, E. A. (1998). The evidence on class size. *Earning and Learning: How Schools Matter*, pages 131–68.
- Hanushek, E. A. (1999). Some findings from an independent investigation of the tennessee star experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis*, 21(2):143–163.
- Hanushek, E. A. (2002). Publicly provided education. In Auerbach, A. J. and Feldstein, M., editors, *Handbook of Public Economics*, volume 4 of *Handbook of Public Economics*, chapter 30, pages 2045–2141. Elsevier.
- Hanushek, E. A. (2013). Economic growth in developing countries: The role of human capital. *Economics of Education Review*, 37:204–212.
- Hanushek, E. A., Woessmann, L., and Zhang, L. (2011). General ducation, vocational education, and labor-market outcomes over the life-cycle. NBER Working Paper 17504, National Bureau of Economic Research.
- Harmon, C., Oosterbeek, H., and Walker, I. (2003). The returns to education: Microeconomics. *Journal of Economic Surveys*, 17(2):115–156.

- Harmon, C. and Walker, I. (1999). The marginal and average returns to schooling in the UK. *European Economic Review*, 43(4):879–887.
- Harmon, C. and Walker, I. (2000). The returns to the quantity and quality of education: Evidence for men in England and Wales. *Economica*, 67(265):19–35.
- Hartog, J. and Oosterbeek, H. (1998). Health, wealth and happiness: why pursue a higher education? *Economics of Education Review*, 17(3):245–256.
- Hartog, J., Pereira, P. T., and Vieira, J. A. (2001). Changing returns to education in portugal during the 1980s and early 1990s: OLS and quantile regression estimators. *Applied Economics*, 33(8):1021–1037.
- Hausman, J. A. and Taylor, W. E. (1981). Panel data and unobservable individual effects. *Journal of Econometrics*, 16(1):155–155.
- Heckman, J. and Polachek, S. (1974). Empirical evidence on the functional form of the earnings-schooling relationship. *Journal of the American Statistical Association*, 69(346):350–354.
- Heckman, J. J., Lochner, L. J., and Todd, P. E. (2006). Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. volume 1 of *Handbook of the Economics of Education*, chapter 7, pages 307–458. Elsevier.
- Heckman, J. J. and Masterov, D. V. (2007). The productivity argument for investing in young children. *Review of Agricultural Economics*, 29(3):446–493.
- Heckman, J. J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–86.
- Heckman, J. J. and Smith, J. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *Review of Economic Studies*, 64(4):487–535.

- Henderson, D. J., Polachek, S. W., and Wang, L. (2011). Heterogeneity in schooling rates of return. *Economics of Education Review*, 30(6):1202–1214.
- Heywood, J. S. and Parent, D. (2012). Performance pay and the white-black wage gap. *Journal of Labor Economics*, 30(2):249–290.
- Hijzen, A., Upward, R., and Wright, P. W. (2010). The income losses of displaced workers. *Journal of Human Resources*, 45(1):243–269.
- Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics*, 115(4):1239–1285.
- Hungerford, T. and Solon, G. (1987). Sheepskin effects in the returns to education. *The Review of Economics and Statistics*, 69(1):175–77.
- Imbens, G. W. (2010). Better late than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature*, 48(2):399–423.
- Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–75.
- Jackson, E. and Page, M. E. (2013). Estimating the distributional effects of education reforms: A look at Project STAR. *Economics of Education Review*, 32(C):92–103.
- Jacob, B. A. and Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, 1(3):33–58.
- Jacobson, L. S., LaLonde, R. J., and Sullivan, D. G. (1993). Earnings losses of displaced workers. *American Economic Review*, 83(4):685–709.
- Kettunen, J. (1997). Education and unemployment duration. *Economics of Education Review*, 16(2):163–170.
- Kletzer, L. G. and Fairlie, R. W. (2003). The long-term costs of job displacement for young adult workers. *Industrial and Labor Relations Review*, 56(4):682–698.

- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.
- Koenker, R. and Hallock, K. (2001). Quantile regression: An introduction. *Journal of Economic Perspectives*, 15(4):43–56.
- Konstantopoulos, S. (2008). Do small classes reduce the achievement gap between low and high achievers? Evidence from Project STAR. *The Elementary School Journal*, 108(4):275–291.
- Koop, G. and Tobias, J. L. (2004). Learning about heterogeneity in returns to schooling. *Journal of Applied Econometrics*, 19(7):827–849.
- Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, 114(2):497–532.
- Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, 113(485):F34–F63.
- Krueger, A. B. and Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project star. *The Economic Journal*, 111(468):1–28.
- Lamarche, C. (2011). Measuring the incentives to learn in Colombia using new quantile regression approaches. *Journal of Development Economics*, 96(2):278–288.
- Lazear, E. P. (2001). Educational production. *The Quarterly Journal of Economics*, 116(3):777–803.
- Lazear, E. P. (2009). Firm-specific human capital: A skill-weights approach. *The Journal of Political Economy*, 117(5):914–940.
- Lefgren, L. (2004). Educational peer effects and the Chicago public schools. *Journal of Urban Economics*, 56(2):169–191.

- Lemieux, T. (2008). The changing nature of wage inequality. *Journal of Population Economics*, 21(1):21–48.
- Machado, J. A. and Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20(4):445–465.
- Maclean, J. C., Webber, D. A., and Marti, J. (2014). An application of unconditional quantile regression to cigarette taxes. *Journal of Policy Analysis and Management*, 33(1):188–210.
- Martins, P. S. and Pereira, P. T. (2004). Does education reduce wage inequality? Quantile regression evidence from 16 countries. *Labour Economics*, 11(3):355–371.
- Maynard, A. and Qiu, J. (2009). Public insurance and private savings: Who is affected and by how much? *Journal of Applied Econometrics*, 24(2):282–308.
- Melly, B. (2006). Estimation of counterfactual distributions using quantile regression. *Review of Labor Economics*, 68(4):543–572.
- Mincer, J. (1991). Education and unemployment. NBER Working Paper 3838, National Bureau of Economic Research.
- Mincer, J. A. (1974). *Schooling, Experience, and Earnings*. NBER Books. National Bureau of Economic Research.
- Monks, J. and Pizer, S. D. (1998). Trends in voluntary and involuntary job turnover. *Industrial Relations*, 37(4):440–459.
- Mosteller, F., Light, R. J., and Sachs, J. A. (1996). Sustained inquiry in education: Lessons from skill grouping and class size. *Harvard Educational Review*, 66(4):797–843.
- Müller, S. (2013). Teacher experience and the class size effect—Experimental evidence. *Journal of Public Economics*, 98(C):44–52.

- Müller, S. (2015). Works councils and labour productivity: Looking beyond the mean. *British Journal of Industrial Relations*, 53(2):308–325.
- Mwabu, G. and Schultz, T. P. (1996). Education returns across quantiles of the wage function: Alternative explanations for returns to education by race in South Africa. *American Economic Review*, 86(2):335–39.
- Neal, D. (1995). Industry-specific human capital: Evidence from displaced workers. *Journal of Labor Economics*, 13(4):653–77.
- Nye, B., Hedges, L. V., and Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis*, 21(2):127–142.
- OECD (2012). Education today 2013: The OECD perspective. Technical report, Organisation for Economic Co-operation and Development (OECD).
- OECD (2013). Education at a glance 2013. Technical report, Organisation for Economic Co-operation and Development (OECD).
- Oreopoulos, P. and Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. *Journal of Economic Perspectives*, 25(1):159–184.
- Park, J. H. (1994). Returns to schooling: A peculiar deviation from linearity. Working Paper 714, Princeton University, Department of Economics, Industrial Relations Section.
- Pischke, J.-S. (2007). The impact of length of the school year on student performance and earnings: Evidence from the German short school years. *The Economic Journal*, 117(523):1216–1242.
- Polsky, D. (1999). Changing consequences of job separation in the United States. *Industrial and Labor Relations Review*, 52(4):565–580.

- Powell, J. L. (1983). The asymptotic normality of two-stage least absolute deviations estimators. *Econometrica*, 51(5):1569–1575.
- Rosenbaum, J. E. and Rosenbaum, J. (2013). Beyond ba blinders: Lessons from occupational colleges and certificate programs for nontraditional students. *Journal of Economic Perspectives*, 27(2):153–172.
- Ruhm, C. J. (1991). Are workers permanently scarred by job displacements? *American Economic Review*, 81(1):319–324.
- Rupietta, C. and Backes-Gellner, U. (2012). High quality workplace training and innovation in highly developed countries. Economics of Education Working Paper Series 0074, University of Zurich, Institute for Strategy and Business Economics (ISU).
- Sakellariou, C. (2012). Unconditional quantile regressions, wage growth and inequality in the Philippines, 2001–2006: The contribution of covariates. *Applied Economics*, 44(29):3815–3830.
- Saniter, N. (2012). Estimating heterogeneous returns to education in Germany via conditional heteroskedasticity. IZA Discussion Papers 6813, Institute for the Study of Labor (IZA).
- Santos Silva, J. M. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658.
- Santos Silva, J. M. and Tenreyro, S. (2011). Further simulation evidence on the performance of the poisson pseudo-maximum likelihood estimator. *Economics Letters*, 112(2):220–222.
- Schanzenbach, D. W. (2006). What have researchers learned from Project STAR? *Brookings Papers on Education Policy*, pages 205–228.
- Schultz, T. W. (1961). Investment in human capital. *American Economic Review*, 51(1):1–17.

- Spence, A. M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374.
- Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.
- Stevens, A. H. (1997). Persistent effects of job displacement: The importance of multiple job losses. *Journal of Labor Economics*, 15(1):165–88.
- Stueber, H. and Beissinger, T. (2012). Does downward nominal wage rigidity dampen wage increases? *European Economic Review*, 56(4):870–887.
- Tang, Y. and Long, W. (2013). Gender earnings disparity and discrimination in urban china: Unconditional quantile regression. *African Journal of Science, Technology, Innovation and Development*, 5(3):202–212.
- Todd, P. E. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485):F3–F33.
- Trostel, P., Walker, I., and Woolley, P. (2002). Estimates of the economic return to schooling for 28 countries. *Labour Economics*, 9(1):1–16.
- Tuor, S. N. and Backes-Gellner, U. (2010). Risk-return trade-offs to different educational paths: vocational, academic and mixed. *International Journal of Manpower*, 31(5):495–519.
- Vandenbussche, J., Aghion, P., and Meghir, C. (2006). Growth, distance to frontier and composition of human capital. *Journal of Economic Growth*, 11(2):97–127.
- von Wachter, T., Song, J., and Manchester, J. (2008). Long-term earnings losses due to job separation during the 1982 recession: An analysis using longitudinal administrative data from 1974 to 2004. Discussion Paper 0708-16, Columbia University, Department of Economics.

- Wang, L. (2013). How does education affect the earnings distribution in urban China? *Oxford Bulletin of Economics and Statistics*, 75(3):435–454.
- Weber, B. and Wolter, S. (1999). Wages and human capital: Evidence from Switzerland. *ETLA: The Research Institute of the Finnish Economy*.
- Wehby, G. L., Murray, J. C., Castilla, E. E., Lopez-Camelo, J. S., and Ohsfeldt, R. L. (2009). Quantile effects of prenatal care utilization on birth weight in argentina. *Health Economics*, 18(11):1307–1321.
- White, R. (2010). Long-run wage and earnings losses of displaced workers. *Applied Economics*, 42(14):1845–1856.
- Willis, R. J. and Rosen, S. (1979). Education and self-selection. *The Journal of Political Economy*, 87(5):7–36.
- Woessmann, L. (2005). Educational production in Europe. *Economic Policy*, 20(43):445–504.
- Woessmann, L. and West, M. (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review*, 50(3):695–736.
- Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, 90(1):77–97.
- Word, E., Johnston, J., Bain, H. P., Fulton, B., Zaharias, J. B., Achilles, C. M., Lintz, M. N., Folger, J., and Breda, C. (1990). The state of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Final summary report 1985-1990. *Nashville: Tennessee State Department of Education*.
- Zwick, T. (2012). Earnings losses after non-employment increase with age. *Schmalenbach Business Review*, 64(1):2–19.